

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 30, 2003, 13:16:08 ; Search time 72 Seconds
(without alignments)
158.727 Million cell updates/sec

Title: US-09-856-681-4

Perfect score: 376
Sequence: 1 PPPAPQRVDSIQVHSSQSPG.....PPKPSFADLSTSMKPNDACT 72

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: A.GeneSeq_19Jun03:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
- 25: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	376	100.0	72	21	AAV71461 Binding domain of
2	376	100.0	507	22	AA92688 Human protein sequ
3	376	100.0	562	22	AA93444 Human polypeptide,
4	376	100.0	562	22	AA94104 Human protein sequ
5	376	100.0	574	22	AA95317 Human protein sequ
6	376	100.0	699	22	AA95139 Human protein sequ
7	376	100.0	863	22	ABG04066 Novel human diagno
8	376	100.0	974	19	AA64221 Human secreted pro
9	376	100.0	975	22	AA90731 Human CJI45_1 pro

10	376	100.0	1030	21	AAV71460 Human semaphorin 6
11	163.5	43.5	451	22	AA94239 Human protein sequ
12	163.5	43.5	464	22	AA94296 Human protein sequ
13	163.5	43.5	474	24	ABU1724 Human MDPF polypep
14	163.5	43.5	1017	23	AA679413 CADHP-2, Incyte ID
15	163.5	43.5	1032	23	ABG79175 Human semaphorin-1
16	163.5	43.5	1035	23	ABG79173 Human semaphorin 6
17	163.5	43.5	1086	22	AA663213 Amino acid sequenc
18	163.5	43.5	1088	22	ABG79177 Human semaphorin-1
19	86	23.9	873	24	ABG74448 Rat germinal centr
20	80.5	21.4	374	22	AA52309 Murine zyxine frag
21	80.5	21.4	564	22	AA52303 Mouse tschaemic co
22	80.5	21.4	564	23	AB57260 Human polypeptide
23	77.5	20.6	693	23	AB69529 Herbicidally activ
24	77.5	20.6	1111	23	AB69387 Drosophila melanog
25	75.5	20.1	735	22	AB66333 Propionibacterium
26	75.5	20.1	745	22	AB668321 Drosophila melanog
27	75	19.9	95	22	AA46062 Human protein SEQ
28	74.5	19.8	449	22	AB60133 Human KHS2 protein
29	74.5	19.7	1475	22	ABH71451 Human protein SEQ
30	74	19.7	894	20	AAV5935 Human KHS2 protein
31	74	19.7	894	20	AAV5935 Human protein SEQ
32	74	19.7	894	20	AAV5935 Human protein SEQ
33	74	19.7	930	22	AAV79329 Human protein SEQ
34	74	19.7	930	22	AAV7930 Human protein SEQ
35	73.5	19.5	425	22	ABG22456 Novel human diagno
36	73.5	19.5	960	22	ABG20323 Novel human diagno
37	73.5	19.5	960	22	ABG20886 Human liver peptid
38	73.5	19.5	2665	22	ABG48316 Peptide #949 encod
39	73.5	19.5	2665	22	ABB28314 Human brain expres
40	73.5	19.5	2665	22	ABB3340 Human bone marrow
41	73.5	19.5	2665	22	ABBI18950 Peptide #967 encod
42	73.5	19.5	2665	22	AA654270 Human peptide #967
43	73.5	19.5	2665	22	AA66665 Peptide #987 encod
44	73.5	19.5	2665	22	AA614533
45	73.5	19.5	2665	22	AA66950

ALIGNMENTS

RESULT 1	AAV71461	AAV71461 standard; peptide; 72 AA.
XX	AAV71461	
AC	AAV71461	
XX	04-OCT-2000	(first entry)
DT		
XX		
DE		Binding domain of human semaphorin 6A-1.
XX		
KW		Human: semaphorin 6A-1; (HSA)SEMA6A-1; neuronal development; apoptosis;
KW		neural regeneration; Ema/VASP protein family; Immunomodulatory;
KW		gene therapy; diagnostic agent; therapeutic agent; differentiation;
KW		cytoskeletal stabilisation; plasticity.
XX		
OS		Homo sapiens.
XX		
FT	Key	Location/Qualifiers
FT	Binding-site	51..56
FT		/note="Specific binding motif for members of
FT		Ema/VASP protein family, especially Evt1"
XX		
PN		
XX		
XX		
PD	02-JUN-2000.	
XX		
PF	26-NOV-1999;	99WO-EP09215.
XX		
PR	26-NOV-1998;	98EP-0122441.
XX		
PA		(PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
XX		

PI Behl C, Klostermann A;
 XX
 DR WPI: 2000-40065/34.
 DR N-PSDB; AAD01234.
 XX
 PT Nucleic acid coding for human semaphorin 6A-1 used as diagnostic agent,
 PT therapeutic agent, for modulating immune system, in gene therapy or for
 PT effecting differentiation, cytoskeletal stabilization and/or plasticity
 PT
 PS Disclosure; Page 22; 53pp; English.
 XX
 CC The present sequence is a binding domain of transmembranous
 CC human semaphorin 6A-1 ((HSA)SEMA6A-1) which is involved in neuronal
 CC development and regeneration mechanisms during apoptosis. The binding
 CC domain shows homology to zyxin protein and selectively binds to members
 CC of Fna/VASP protein family, especially Evi. (HSA)SEMA6A-1 is a
 CC member of protein family displaying secreted or transmembrane-based
 CC repulsive guidance cues critically involved in neuronal development.
 CC Expression of (HSA)SEMA6A-1 is highest in embryonic brain and
 CC kidney and moderate in lung. The present sequence is useful as diagnostic
 CC and therapeutic agents, for modulating the immune system, in gene
 CC therapy, for effecting differentiation, cytoskeletal stabilisation
 CC and plasticity.
 XX
 SQ Sequence 72 AA:
 Query Match 100.0%; Score 376; DB 21; Length 72;
 Best Local Similarity 100.0%; Pred. No. 7.4e-35;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PPPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 60
 Db 1 PPPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 60
 QY 61 LSTSMKRPNDACT 72
 Db 61 LSTSMKRPNDACT 72
 XX
 RESULT 2
 AAB92688
 ID AAB92688 standard; Protein: 507 AA.
 XX
 AC AAB92688;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:11073.
 XX
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 XX Claim 8; SEQ ID 11073; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH33628 and
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 507 AA:
 Query Match 100.0%; Score 376; DB 22; Length 507;
 Best Local Similarity 100.0%; Pred. No. 8.3e-34;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PPPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 60
 Db 436 PPPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 495
 QY 61 LSTSMKRPNDACT 72
 Db 496 LSTSMKRPNDACT 507
 XX
 RESULT 3
 AAM93444
 ID AAM93444 standard; Protein: 562 AA.
 XX
 AC AAM93444;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 3088.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 PN EP130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 XX
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX

XX DR WPI; 2001-524255/58.
 XX DR N-PSDB; AAK94365.
 XX PT 830 Primers useful for synthesizing full length cDNA clones and their
 XX use in genetic manipulation -
 XX PS Claim 8; SEQ ID NO 3088; 1380bp + sequence listing; English.
 XX CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX SQ Sequence 562 AA;
 XX Query Match 100.0%; Score 376; DB 22; Length 562;
 XX Best Local Similarity 100.0%; Pred. No. 9.4e-34;
 XX Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 PPAPORVDSIQVHSSQSGQAVTVSRPSTLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 60
 Db 491 PPPPQRVDSIQVHSSQSGQAVTVSRPSTLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 550
 OY 61 LSTSMKRPNDACT 72
 Db 551 LSTSMKRPNDACT 562
 RESULT 4
 ID AAB94104 standard; Protein: 562 AA.
 AC AAB94104;
 XX 26-JUN-2001 (first entry)
 DT XX
 DE Human protein sequence SEQ ID NO:14328.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KM Homo sapiens.
 OS
 XX EPI074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the

PT Full-length cDNAs -
 XX PS Claim 8; SEQ ID 14328; 2537bp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAB92446 to
 CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAB13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX SQ Sequence 562 AA;
 XX Query Match 100.0%; Score 376; DB 22; Length 562;
 XX Best Local Similarity 100.0%; Pred. No. 9.4e-34;
 XX Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 PPAPORVDSIQVHSSQSGQAVTVSRPSTLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 60
 Db 491 PPPPQRVDSIQVHSSQSGQAVTVSRPSTLNAYNSLTRSGLKRTPLKDPVPPKPSFAP 550
 OY 61 LSTSMKRPNDACT 72
 Db 551 LSTSMKRPNDACT 562
 RESULT 5
 ID AAB95317 standard; Protein: 574 AA.
 AC AAB95317;
 XX 26-JUN-2001 (first entry)
 DT XX
 DE Human protein sequence SEQ ID NO:17568.
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KM Homo sapiens.
 OS
 XX EPI074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the

PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 XX WPI: 2001-639362/73.
 DR N-PSDB; AAS68253.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS
 PS Claim 20: SEQ ID No 34425; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 863 AA;
 Query Match 100.0%; Score 376; DB 22; Length 863;
 Best Local Similarity 100.0%; Pred. No. 1.6e-33;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNATNSLTRSGIKRTPSLKPDVPPKPSFAP 60
 Db 792 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNATNSLTRSGIKRTPSLKPDVPPKPSFAP 851
 QY 61 LSTSMKPNDACT 72
 Db 852 LSTSMKPNDACT 863
 DE Human secreted protein from clone CJI45_1.
 XX
 XX Secreted protein; human fetal brain; nutrition; cytokine; stimulant;
 KW cell proliferation; differentiation; immune system; suppressor; ligand;
 KW regulator; hematopoiesis; tissue growth; activin; inhibin; haemostatic;
 KW chemotaxis; chemokinesis; thrombosis; receptor; cadherin; tumour;
 KW anti-inflammatory.
 XX
 XX Homo sapiens.
 OS
 XX
 XX W09827205-A2.
 PN

XX
 PD 25-JUN-1998.
 XX
 PE 17-DEC-1997; 97WO-US23330.
 XX
 PR 16-DEC-1997; 97US-0991872.
 PR 18-DEC-1996; 96US-0769192.
 PR 13-JAN-1997; 97US-0783401.
 XX
 PA (GENY) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racie LA, Spaulding V, Treacy M;
 XX
 DR WPI: 1998-362774/31.
 DR N-PSDB; AAV44295.
 XX
 PT New polynucleotides and secreted proteins - obtained from human
 PT fetal brain, human adult testes, human adult brain and human adult
 PT salivary gland cDNA libraries
 PS
 PS Claim 17j: Page 71-74; 110pp; English.
 XX
 CC This sequence represents a novel secreted protein from clone CJI45_1
 CC isolated from a human fetal brain cDNA library. This protein has
 CC applications for nutritional use, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating or
 CC suppressing activity, hematopoiesis regulating activity, tissue growth
 CC activity activin/inhibin activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombotic activity, receptor/ligand activity,
 CC anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
 CC tumour inhibition activity and other activities.
 XX
 SQ Sequence 974 AA;
 Query Match 100.0%; Score 376; DB 19; Length 974;
 Best Local Similarity 100.0%; Pred. No. 1.9e-33;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNATNSLTRSGIKRTPSLKPDVPPKPSFAP 60
 Db 903 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNATNSLTRSGIKRTPSLKPDVPPKPSFAP 962
 QY 61 LSTSMKPNDACT 72
 Db 963 LSTSMKPNDACT 974
 DE Human CJI45_1 protein sequence SEQ ID 161.
 XX
 XX Human; secreted protein; nutrient; cytokine modulator; proliferation;
 KW differentiation; immune system modulator; tissue growth; chemotactic;
 KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
 KW hematopoiesis.
 XX
 XX Homo sapiens.
 OS
 XX
 XX W0200119988-A1.
 XX
 PD 22-MAR-2001.
 PF 14-SEP-2000; 2000WO-US25135.
 PR 17-SEP-1999; 99US-0398829.
 XX
 XX
 XX

(GEMV) GENETICS INST INC.

PA XX Jacobus K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
 XX WPI: 2001-244801/25.
 DR N-PSDB; AAF98469.

XX XX Isolated nucleic acids encoding polypeptides, useful for modulating
 PT e.g. cytokine and cell proliferation/differentiation activity, the
 PT immune system and hematopoiesis regulating activity -
 XX
 PS Disclosure; Page 487-490; 557pp; English.

XX XX Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
 CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
 CC tissue types, and may be used in the prevention, treatment and diagnosis
 CC of diseases associated with inappropriate protein expression. The
 CC polypeptides and nucleic acids may be used as nutrients or to modulate
 CC cytokine and cell proliferation/differentiation activity and may also be
 CC involved in modulation of the immune system. The cDNA sequences,
 CC proteins, their agonists and/or antagonists exhibit haematopoiesis
 CC regulating activity; tissue growth activity; activin/inhibin activity;
 CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
 CC activity; receptor/ligand activity; anti-inflammatory activity; and/or
 CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
 CC tumour inhibition activity. Included in the invention are probes
 CC represented in AAF98490 - AAF98572 which are specific for the cDNA clones
 CC encoding the secreted proteins.
 CC
 XX Sequence 975 AA;

SO Query Match 100.0%; Score 376; DB 22; Length 975;
 Best Local Similarity 100.0%; Pred. No. 1.9e-33;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 60
 DB 904 PPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 963
 QY 61 LSTSMKPNDACT 72
 DB 964 LSTSMKPNDACT 975

RESULT 10
 AAY71460
 ID AAY71460 standard; Protein; 1030 AA.
 XX
 AC AAY71460;
 XX
 DT 04-OCT-2000 (first entry)
 XX
 DE Human semaphorin 6A-1.
 XX
 KW Human: semaphorin 6A-1; (HSA)SEMA6A-1; neuronal development; apoptosis;
 KW neuronal regeneration; Ena/VASP protein family; immunomodulatory;
 KW gene therapy; diagnostic agent; therapeutic agent; differentiation;
 KW cytoskeletal stabilisation; plasticity.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Binding-site 959..1030
 FT /note= "Zyxin-like domain that selectively binds to
 FT members of Ena/VASP protein family, especially Evi1"
 FT 957..961
 FT Binding-site
 FT /note= "Specific binding motif for members of
 FT Ena/VASP protein family, especially Evi1"
 FT 1009..1014
 FT Binding-site
 FT /note= "Specific binding motif for members of
 FT Ena/VASP protein family, especially Evi1"
 FT
 XX

PN W0200031252-A1.
 XX
 PD 02-JUN-2000.
 XX
 PF 26-NOV-1999; 99WO-EP09215.
 XX
 PR 26-NOV-1998; 98EP-0122441.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX
 PI Behl C, Klostermann A;
 XX
 DR WPI: 2000-400065/34.
 DR N-PSDB; AAD01233.

XX XX Nucleic acid coding for human semaphorin 6A-1 used as diagnostic agent,
 PT therapeutic agent, for modulating immune system, in gene therapy or for
 PT effecting differentiation, cytoskeletal stabilization and/or plasticity
 PT
 XX
 PS Example 1; Page 29-33; 53pp; English.

XX XX The present sequence is a transmembranous human semaphorin
 CC 6A-1 ((HSA)SEMA6A-1) which is involved in neuronal development and
 CC regeneration mechanisms during apoptosis. Semaphorin is a family of
 CC proteins displaying secreted or transmembrane-based repulsive guidance
 CC cues critically involved in neuronal development. The present sequence
 CC was isolated from human 1-ZAP Express cDNA library which was screened
 CC using a PCR fragment amplified from human neuroblastoma cell line
 CC SK-N-MC cDNA. The (HSA)SEMA6A-1 protein contains a zyxin-like domain
 CC that selectively binds to members of Ena/VASP protein family especially
 CC Evi1. Expression of (HSA)SEMA6A-1 is highest in embryonic brain and
 CC kidney and moderate in lung. The present sequence is useful as diagnostic
 CC and therapeutic agents, for modulating the immune system, in gene
 CC therapy, for effecting differentiation, cytoskeletal stabilisation
 CC and plasticity.
 CC
 XX Sequence 1030 AA;

SO Query Match 100.0%; Score 376; DB 21; Length 1030;
 Best Local Similarity 100.0%; Pred. No. 2e-33;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 60
 DB 959 PPAPQVRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 1018
 QY 61 LSTSMKPNDACT 72
 DB 1019 LSTSMKPNDACT 1030

RESULT 11
 AAB94239
 ID AAB94239 standard; Protein; 451 AA.
 XX
 AC AAB94239;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:14623.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 8; SEQ ID 14623; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 451 AA;
Query Match 43.5%; Score 163.5; DB 22; Length 451;
Best Local Similarity 50.7%; Pred. No. 5.9e-10;
Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;
OY 1 PPAPQRVDSIQVHSSQPSQAVTVSRQPSLNAVNSLT-----RSGLKRPSTLKPDPVP 54
DB 380 PTPGAKVDYID-----GTPVSVHLQPSLSRSSSTYSNGTLPTGLKRPSTLKPDPVP 432
OY 55 KPSPAPLSTSMKP 67
DB 433 KPSEVPQTPSVRP 445
RESULT 12
AAB94296
ID AAB94296 standard; Protein: 464 AA.
XX
AC AAB94296;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14749.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX
XX EPI074617-A2.
XX
XX

PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 8; SEQ ID 14749; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 464 AA;
Query Match 43.5%; Score 163.5; DB 22; Length 464;
Best Local Similarity 50.7%; Pred. No. 6.1e-10;
Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;
OY 1 PPAPQRVDSIQVHSSQPSQAVTVSRQPSLNAVNSLT-----RSGLKRPSTLKPDPVP 54
DB 393 PTPGAKVDYID-----GTPVSVHLQPSLSRSSSTYSNGTLPTGLKRPSTLKPDPVP 445
OY 55 KPSPAPLSTSMKP 67
DB 446 KPSEVPQTPSVRP 458
RESULT 13
ABU11724
ID ABU11724 standard; Protein: 474 AA.
XX
AC ABU11724;
XX
DT 13-FEB-2003 (first entry)
XX
DE Human MDDT polypeptide SEQ ID 671.
XX
XX MDDT: human; disease detection and treatment molecule polypeptide;
XX

ET /note= "Potentially glycosylated"
 ET 459
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 461
 ET Modified-site
 ET /note= "Potentially glycosylated"
 ET 513
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 520
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 576
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 602..630
 ET Domain
 ET /label= "Transmembrane domain"
 ET /note= "Identified by TMAP, N-terminal domain is cytoplasmic"
 ET 650
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 678
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 687
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 688
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 734
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 736
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 745
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 749
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 776
 ET Modified-site
 ET /note= "Potentially glycosylated"
 ET 782
 ET Modified-site
 ET /note= "Potentially glycosylated"
 ET 808
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 809
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 822
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 858
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 886
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 900
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 911
 ET Modified-site
 ET /note= "Potentially glycosylated"
 ET 913
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 978
 ET Modified-site
 ET /note= "Potentially glycosylated"
 ET 988..1017
 ET Binding-site
 ET /label= "Semaphorin protein precursor_receptor"
 ET /note= "Identified by BLAST-PRODOM"
 ET 991
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET 1008
 ET Modified-site
 ET /note= "Potentially phosphorylated"
 ET
 PN WO200259312-A2.
 PD 01-AUG-2002.
 XX 18-DEC-2001; 2001MO-US49206.
 XX 18-DEC-2000; 2000US-256542P.
 PR 22-DEC-2000; 2000US-259604P.
 PR 05-JAN-2001; 2001US-260101P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX

PI Duggan BM, Xu Y, Lee EA, Lee S, Lu DAM, Warren RA, Yue H;
 PI Gietzen KU, Honchell CD, Burford N, Baughn MR, Tang TY;
 PI Hillman JL, Gandhi AR, Kalliock DA, Bandman O, Grail RC, Wallia NK;
 PI Lu Y, Rankumar J, Yao MG, Lal PG;
 XX
 DR WPI; 2002-590826/63.
 DR N-PSDB; ABA00055.
 XX
 PT New human cell adhesion proteins (CADHP) useful for treating,
 PT diagnosing and preventing diseases or conditions associated with the
 PT aberrant CADHP expression e.g. cancer, acquired immunodeficiency
 PT syndrome, Alzheimer's disease and epilepsy
 PS
 PS Claim 1; Page 115-17; 149pp; English.
 XX
 CC The sequences given in AAG79412-21 are novel human cell adhesion
 CC proteins (CADHP). The CADHP polypeptides and polynucleotides are useful
 CC in treating, diagnosing and preventing diseases or conditions associated
 CC with the decreased expression or overexpression of CADHP, e.g. immune
 CC system (acquired immunodeficiency syndrome, thymic dysplasia),
 CC neurological (Alzheimer's disease, Parkinson's disease, epilepsy),
 CC developmental (renal tubular acidosis, congenital glaucoma) and cell
 CC proliferative (cancer, atherosclerosis) disorders. They are also useful
 CC in assessing the effects of exogenous compounds on the expression of
 CC nucleic acid and amino acid sequences of CADHP. The CADHP or its
 CC fragments are useful in screening compounds for effectiveness as
 CC agonist or antagonist of the polypeptides, or in altering the
 CC expression of the target polynucleotide and compounds that specifically
 CC bind to or modulate the activity of the polypeptide. This protein
 CC shows homology to mouse semaphorin VIa.
 CC
 SQ Sequence 1017 AA;
 Query Match 43.5%; Score 163.5; DB 23; Length 1017;
 Best Local Similarity 50.7%; Pred. No. 1.6e-09;
 Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;
 QY 1 PPPAPQRNDSTGVHSSQSGCAVTVSRQPSLNAYNSLT-----RSGLKRTPSLKDVP 54
 Db 946 PPTGAKVDYID-----GTPVSHLOPSLSRQSSYTSNGTLPTGTLKRTPSLKDVP 998
 QY 55 KPSEAPLSTSMKP 67
 Db 999 KPSEVPTQTSVMP 1011
 RESULT 15
 ABG79175
 ID ABG79175 standard; Protein; 1032 AA.
 AC ABG79175;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Human semaphorin-like protein #3.
 XX
 KW Human; MOVX; neurodegenerative disease; Alzheimer's disease; anxiety;
 KW Parkinson's disease; Huntington's disease; neurological disorder;
 KW schizophrenia; manic depression; mental retardation; angina pectoris;
 KW cardiovascular disease; acute heart failure; myocardial infarction;
 KW muscular disease; muscular disorder; retinal disease; photoreception;
 KW deafness; keratinisation disorder; cancer; ovarian cancer; melanoma;
 KW immunological disorder; inflammatory disease; immune disease; diabetes;
 KW bacterial infection; fungal infection; protozoal infection; obesity;
 KW viral infection; reproductive system disorder; metabolic disturbance;
 KW anorexia; wasting disorder; chronic disease; infectious disease;
 KW dyslipidaemia; cub; sushi; myelin; von willebrand factor; kielin;
 KW semaphorin; serine/threonine protein kinase; TGF-beta binding;
 KW mas proto-oncogene; ribonuclease pancreatic precursor; aminotransferase;
 KW tollid-like 2; cysteine sulfonic acid decarboxylase.
 XX
 OS Homo sapiens.
 XX

Job time : 74 secs

```

PN WO200264791-A2.
XX
XX 22-AUG-2002.
XX
XX 10-DEC-2001; 2001WO-US48369.
XX
XX 08-DEC-2000; 2000US-254329P.
PR 14-DEC-2000; 2000US-255648P.
PR 15-MAY-2001; 2001US-291037P.
PR 08-JUN-2001; 2001US-297173P.
PR 08-JUN-2001; 2001US-309258P.
PR 29-AUG-2001; 2001US-315639P.
PR 01-OCT-2001; 2001US-326393P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Alabrook JP, Anderson DW, Burgess CE, Boldog FL, Casman SJ,
PI Colman SD, Edinger SR, Ellerman K, Gerlach V, Gorman L, Grosse WM;
PI Guo X, Herrmann JL, Kekuda R, Lepley DM, Li L, MacDougall JR;
PI Millet I, Pena CE, Peyman JA, Rastelli L, Rieger DK, Shinkets RA;
PI Smithson G, Spytek KA, Stone DJ, Tchernev VT, Vernet CAM, Voss EZ;
PI Zernhusen BD, Zhong H, Zhong M;
XX
XX WPI; 2002-643486/69.
DR N-PSDB; ABS64382.
XX
XX New NOVX polypeptides and polynucleotides useful for treating or
PT preventing e.g. neurodegenerative diseases, neurological disorders,
PT cardiovascular diseases, muscular diseases and disorders, or
PT immunological diseases
XX
XX Claim 1; Page 55; 299pp; English.
XX
XX The present invention relates to new NOVX polypeptides. The polypeptides,
CC polynucleotides and antibodies are useful in the manufacture of a
CC medicament for treating or preventing neurodegenerative diseases (e.g.
CC Alzheimer's disease, Parkinson's disease, or Huntington's disease),
CC neurological disorders (e.g. anxiety, schizophrenia, manic depression or
CC mental retardation), cardiovascular disease (e.g. acute heart failure,
CC angina pectoris or myocardial infarction), muscular diseases and
CC disorders, retinal diseases (including those involving photoreception,
CC deafness and keratinisation disorders), cancer (e.g. ovarian cancer or
CC melanoma), immunological disorders, inflammatory and immune diseases,
CC bacterial, fungal, protozoal and viral infections, and reproductive
CC system disorders. The proteins of the invention may be used to screen
CC drugs or compounds that modulate the NOVX protein activity or expression,
CC as well as to treat disorders characterised by insufficient or excessive
CC production of NOVX protein or protein forms that have decreased or
CC aberrant activity compared to NOVX wild type protein, such as diabetes,
CC obesity, metabolic disturbances associated with obesity, anorexia and
CC wasting disorders associated with chronic diseases and various cancers,
CC infectious diseases and various dyslipidaemias. The nucleic acid
CC sequences of the invention may be used in chromosome mapping,
CC identifying an individual from minute biological samples (tissue typing),
CC and in forensic identification of a biological sample. The present
CC amino acid sequence represents a NOVX protein of the invention.
XX
XX
SQ Sequence 1032 AA;

```

```

Query Match 43.5%; Score 163.5; DB 23; Length 1032;
Best local Similarity 50.7%; Pred. No. 1.7e-09;
Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;

```

```

QY 1 PPPAPQRYDSIQVHSSQSPGQAVYRSRPSLNAVNSLT-----RSGLKRPISLKPDPVP 54
DB 961 PTPGAKVDYIQ-----GTPSVYHLPISLROSSSTYSNGTLPRTGLKRTPSLKPDPVP 1013
QY 55 KPSEAPISMKP 67
DB 1014 KPSEFVPQPPSVRP 1026

```

Search completed: September 30, 2003, 16:32:09

US-09-252-991A-32313
; Sequence 32313, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32313
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32313

Query Match 18.4%; Score 69; DB 4; Length 400;
Best Local Similarity 30.1%; Pred. No. 8.2;
Matches 25; Conservative 9; Mismatches 25; Indels 24; Gaps 5;
DB 113 PPAPRV-----DSIQVHSSQPSGQAVTSCPSLNAYNSLTRSGIKRTPS 47
DB 48 LKPDVPPKPSFAPLSTSMKPND 70
DB 166 NRPOPPATSESPMS---KPDPA 185

RESULT 7
US-09-196-270-6
; Sequence 6, Application US/09196270
; Patent No. 6500636
; GENERAL INFORMATION:
; APPLICANT: Hecht, Peter
; APPLICANT: Madden, Kevin
; APPLICANT: Fink, Gerald
; TITLE OF INVENTION: CHIMERIC PRE-ACTIVATED TRANSCRIPTION
; TITLE OF INVENTION: FACTORS
; FILE REFERENCE: 50078/004002
; CURRENT APPLICATION NUMBER: US/09/196,270
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: 60/066,129
; EARLIER FILING DATE: 1997-11-19
; EARLIER APPLICATION NUMBER: 60/066,308
; EARLIER FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: 60/066,462
; EARLIER FILING DATE: 1997-11-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-196-270-6

Query Match 18.2%; Score 68.5; DB 4; Length 625;
Best Local Similarity 25.8%; Pred. No. 16;
Matches 31; Conservative 13; Mismatches 27; Indels 49; Gaps 6;
DB 1 PPAPRVDSIQVHSS-----QPSQAVT-----VSRO- 28
DB 418 PPNABS-YQVGGSSISATANTATYVPLAKYPTGPTHTPLHSNTAGVGNROS 476
DB 29 -----PSLNAYNSLTRSGIKRTPSLKDVPPKPSFALS--TSKKPN-----DACT 72
DB 477 QYAMPHPVRAAPSYSSGCSITLPLQRIKPLPRTMAGTSLKPMWERSLNOKSCT 536

RESULT 8
US-08-560-005-5
; Sequence 5, Application US/08560005
; Patent No. 6001354
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Jefferson, Anne Bennett
; APPLICANT: Majerus, Philip W.
; TITLE OF INVENTION: No. 6001354e1 Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,005
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-0624000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1149 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..1149
; OTHER INFORMATION: /note="51c"
US-08-560-005-5

Query Match 18.1%; Score 68; DB 3; Length 1149;
Best Local Similarity 29.9%; Pred. No. 40;
Matches 23; Conservative 12; Mismatches 36; Indels 6; Gaps 3;
DB 2 PPAPRV---DSIQVHSSQPSG--QAVTVSROPSLNAYNSLTRSGIKRTP-SIKPDVPPK 55
DB 880 PPAPRAAPREELPRLKPEGAPEEGVAAPPKNSFNNPAYVLEGVPHQLPEPPS 939
DB 56 PSFAPLSTSMKPDACT 72
DB 940 PARAPVPSATKNKVALIT 956

RESULT 9
US-09-418-540-5
; Sequence 5, Application US/09418540
; Patent No. 6296848
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; APPLICANT: Williams, Lewis T.
; APPLICANT: Majerus, Philip W.
; APPLICANT: Jefferson, Anne Bennett
; TITLE OF INVENTION: No. 6296848e1 Grb2 Associating Protein and Nucleic
; TITLE OF INVENTION: Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: #1.0, Version #1.25
APPLICATION NUMBER: US/09/418, 540
FILING DATE: 14-OCT-1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/560,005
FILING DATE: 17-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0624000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 1..1149
OTHER INFORMATION: /note= "51c"
US-09-418-540-5

Query Match
Best Local Similarity 18.1%; Score 68; DB 3; Length 1149;
Matches 23; Conservative 12; Mismatches 36; Indels 6; Gaps 3;

QY 2 PPAPQGV--DSIQVHSSQPSG--QAVTVSRQPSINAYNSLTRSGIKRTP-SLKRDVPPK 55
DB 880 PPAPRAAPREEPILPRIKPBGAPBEGVAAPPKNNSFNPPAYVYLEGVPHQLPPEPPS 939
QY 56 PSFAPLSTSMKPNDACT 72
DB 940 PARAPVPSATKNKVAIT 956

RESULT 10
US-09-969-528-5
; Sequence 5, Application US/09969528
; Patent No. 6472197
; GENERAL INFORMATION:
; APPLICANT: Pot, David A.
; Williams, Lewis T.
; Jefferson, Anne Bennett
; Majerus, Philip W.
; TITLE OF INVENTION: No. 6472197el Grp2 Associating Protein and Nucleic
; Acids Encoding Therefor
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/969,528
FILING DATE: 01-Oct-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/560,005
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 2307K-0624000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1149 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 1..1149
OTHER INFORMATION: /note= "51c"
US-09-969-528-5

Query Match
Best Local Similarity 18.1%; Score 68; DB 4; Length 1149;
Matches 23; Conservative 12; Mismatches 36; Indels 6; Gaps 3;

QY 2 PPAPQGV--DSIQVHSSQPSG--QAVTVSRQPSINAYNSLTRSGIKRTP-SLKRDVPPK 55
DB 880 PPAPRAAPREEPILPRIKPBGAPBEGVAAPPKNNSFNPPAYVYLEGVPHQLPPEPPS 939
QY 56 PSFAPLSTSMKPNDACT 72
DB 940 PARAPVPSATKNKVAIT 956

RESULT 11
US-08-252-966B-12
; Sequence 12, Application US/08252966B
; Patent No. 5624818
; GENERAL INFORMATION:
; APPLICANT: Eisenman, Robert N.
; Applicant: Hurlin, Peter J.
; APPLICANT: Ayer, Donald E.
; TITLE OF INVENTION: Regulatory Proteins that Dimerize with
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,966B
; FILING DATE: 01-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: FHCRL7694

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-8100
 TELEFAX: (206) 224-0779
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1253 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DESCRIPTION: translation of msina cDNA; see Figure 23
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Mus musculus
 US-08-252-966B-12

Query Match 18.1%; Score 68; DB 1; Length 1253;
 Best Local Similarity 30.7%; Pred. No. 44;
 Matches 23; Conservative 7; Mismatches 35; Indels 10; Gaps 3;

QY 1 PPAAPQVDSIOVHSSQSGAVTVSRQPS--LNNVNSLTRSGIKRTSLKPDVPPKPSF 58
 Db 238 PQPPQ-----HPSQSSQSAFPAPQAPQPTAKVSKPSQLQAHTPASQOTPLPPY 290
 QY 59 A-PLTSMKPNDACT 72
 Db 291 ASRSPPPVQPHTPVT 305

RESULT 12
 US-08-252-966B-18
 Sequence 18, Application US/08252966B
 Patent No. 5624818
 GENERAL INFORMATION:
 APPLICANT: Eisenman, Robert N.
 APPLICANT: Hurlin, Peter J.
 APPLICANT: Ayer, Donald E.
 TITLE OF INVENTION: Regulatory proteins that dimerize with
 TITLE OF INVENTION: Mad or Max
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessPLLC
 STREET: 1420 Fifth Ave., Suite 2800
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98101-2347
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/252,966B
 FILING DATE: 01-JUN-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Shelton, Dennis K.
 REGISTRATION NUMBER: 26,997.
 REFERENCE/DOCKET NUMBER: FHCRI1694
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-8100
 TELEFAX: (206) 224-0779
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1261 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DESCRIPTION: translation of msina9 cDNA; see Figure 29A, B, C, D
 HYPOTHEICAL: YES
 ORIGINAL SOURCE:

ORGANISM: Mus musculus
 US-08-252-966B-18

Query Match 18.1%; Score 68; DB 1; Length 1261;
 Best Local Similarity 30.7%; Pred. No. 44;
 Matches 23; Conservative 7; Mismatches 35; Indels 10; Gaps 3;

QY 1 PPAAPQVDSIOVHSSQSGAVTVSRQPS--LNNVNSLTRSGIKRTSLKPDVPPKPSF 58
 Db 238 PQPPQ-----HPSQSSQSAFPAPQAPQPTAKVSKPSQLQAHTPASQOTPLPPY 290
 QY 59 A-PLTSMKPNDACT 72
 Db 291 ASRSPPPVQPHTPVT 305

RESULT 13
 US-09-252-991A-30563
 Sequence 30563, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 30563
 LENGTH: 169
 TYPE: PRT
 ORGANISM: pseudomonas aeruginosa
 US-09-252-991A-30563

Query Match 17.8%; Score 67; DB 4; Length 169;
 Best Local Similarity 27.4%; Pred. No. 46;
 Matches 23; Conservative 14; Mismatches 29; Indels 18; Gaps 3;

QY 3 PPAAPQVDSIOVHSSQSGAVTVSRQPSLNNVNSLTRSGIKRTSLKPDVPPKPSF 49
 Db 2 PCSRATVTPSSSSPAPACSCGRPNFRAGSTLPPASRCCRSRSTRSTRTG 61
 QY 50 --PDVPPKPSFAPLSTSMKPNDACT 71
 Db 62 WPTVPVPA--ASTSTRPTRC 82

RESULT 14
 US-08-046-585-5
 Sequence 5, Application US/08046585
 Patent No. 545362
 GENERAL INFORMATION:
 APPLICANT: Iamarco, Kelly
 APPLICANT: Wilson, Angus
 APPLICANT: Herr, Winship
 TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
 TITLE OF INVENTION: HOST CELL FACTOR
 NUMBER OF SEQUENCES: 15
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: USA
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,585
; FILING DATE: 12-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-1/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELETYPE: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-046-585-5

```

```

Query Match      17.8%; Score 67; DB 1; Length 2035;
Best Local Similarity 32.8%; Pred. No. 1e+02; Indels 8; Gaps 3;
Matches 22; Conservative 10; Mismatches 27;

```

```

QY 2 PPAPQVDSIQVHSQPSG-----QAVTVSRQPSLNAYNSLTR--SGIKRTPSLKPDVPP 54
Db 1741 PSTVALLPSTATESLAPSNTEVAPQPVVAVASPAKIQAAATLLEVANGIE-SLGVKPDLP 1799
QY 55 KPSFAPL 61
Db 1800 PPSKAPM 1806

```

```

RESULT 15
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/393,703
; FILING DATE: 24-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-2/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELETYPE: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2035 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-393-703-5

```

```

Query Match      17.8%; Score 67; DB 1; Length 2035;
Best Local Similarity 32.8%; Pred. No. 1e+02; Indels 8; Gaps 3;
Matches 22; Conservative 10; Mismatches 27;

```

```

QY 2 PPAPQVDSIQVHSQPSG-----QAVTVSRQPSLNAYNSLTR--SGIKRTPSLKPDVPP 54
Db 1741 PSTVALLPSTATESLAPSNTEVAPQPVVAVASPAKIQAAATLLEVANGIE-SLGVKPDLP 1799
QY 55 KPSFAPL 61
Db 1800 PPSKAPM 1806

```

```

Search completed: September 30, 2003, 16:36:16
Job time : 43 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 30, 2003, 16:22:28 ; Search time 44 Seconds

(without alignments)
157.367 Million cell updates/sec

Title: US-09-856-681-4

Perfect score: 376

Sequence: 1 PPAPQRVDSIQVHSSQPSG.....PPKSFAPLSTSMKPNDACR 72

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	23.1	961	2 A55380	faciogenital dyspl
2	80.5	21.4	1322	2 A59288	myosin heavy chain
3	77.5	20.6	1111	2 T05646	hypothetical prote
4	77	20.5	175	2 T47463	serine/proline-ric
5	76.5	20.3	744	2 E86255	hypothetical prote
6	75.5	20.1	393	2 T33103	lin-1 protein - Ca
7	75.5	20.1	494	2 A42170	zinc finger protei
8	75.5	20.1	497	2 JG5076	myc-associated zin
9	74	19.7	452	2 S23199	imidazoleglycerol-
10	73.5	19.5	2282	2 T42717	DNA-binding protei
11	73	19.4	867	2 T41308	hypothetical zinc
12	72.5	19.3	628	2 S01955	hypothetical prote
13	72.5	19.3	657	2 B84869	probable SP16 prot
14	72.5	19.3	4957	2 T03455	ALR protein - huma
15	72.5	19.3	5262	2 T03454	ALR protein - huma
16	72	19.1	459	2 A41977	retinoic acid rece
17	71.5	19.0	446	2 A42029	transcription fact
18	71.5	19.0	1522	2 T39371	transcription regu
19	71.5	19.0	2957	2 T33152	hypothetical prote
20	71	18.9	621	2 JG7278	adaptor protein co
21	71	18.9	1150	2 S58775	mypl protein - smu
22	70.5	18.8	468	2 T48615	hypothetical prote
23	70.5	18.8	1420	2 T37781	probable cytoskele
24	70	18.6	719	2 S62466	probable ATP-depen
25	70	18.6	747	2 S35546	ATP-dependent RNA
26	70	18.6	792	2 T26050	hypothetical prote
27	70	18.6	1012	2 I53172	RAE-28 - mouse
28	70	18.6	1201	2 G86441	unknown protein [i
29	69.5	18.5	331	2 B47236	zinc-finger protei

30	69.5	18.5	477	2 A47236	zinc-finger protei
31	69.5	18.5	1125	2 B41206	microtubule-associ
32	69	18.4	3942	2 T42730	Bassoon protein -
33	68.5	18.2	625	2 S48941	regulatory protein
34	68.5	18.2	1106	2 T31742	hypothetical prote
35	68.5	18.2	1188	2 S49915	extensin-like prot
36	68	18.1	428	1 TYH0K	transforming prote
37	68	18.1	530	2 A45690	transactivator EBN
38	68	18.1	613	2 T47975	auxin response fac
39	68	18.1	736	2 T25447	hypothetical prote
40	68	18.1	963	2 T40873	probable transcrip
41	68	18.1	1172	2 T00065	hypothetical prote
42	68	18.1	1219	2 I61713	co-repressor prote
43	68	18.1	1229	2 A56068	co-repressor prote
44	68	18.1	1258	2 JG5765	inositol polyphosp
45	67.5	18.0	429	2 JG4965	elk1 protein - mou

ALIGNMENTS

RESULT 1
A55380
faciogenital dysplasia-associated protein FGDI - human
C/Species: Homo sapiens (man)
C/Date: 10-Feb-1995 #sequence,revision 10-Feb-1995 #text_change 17-Mar-1999
C/Accession: A55380
R/Pastorius, N.G.; Cadle, A.; Loe, L.J.; Porteous, M.F.M.; Schwartz, C.E.; Stevenson
Cell 79, 669-678, 1994
A/Title: Isolation and characterization of the faciogenital dysplasia (Aarskog-Scott
A/Reference number: A55380; MUID:95042764; PMID:7954831
A/Accession: A55380
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-961 <PAS>
A/Cross-references: GB:U11690; NID:9595424; PID:9595425
C/Superfamily: CDC24 homology; plectstrin repeat homology
E/373-561/Domain: CDC24 homology <CD24>

Query Match 23.1%; Score 87; DB 2; Length 961;
Best local similarity 34.8%; Pred. No. 1.2;
Matches 23; Conservative 5; Mismatches 22; Indels 16; Gaps 2;

QY	2 PPAPQRVDSIQVHSSQPSQAVTVSRQPSLNVNSITRSGIKRTPSPKPSFAPL 61
DB	127 PGQPQRL-----RSDPGPPTPTPSQRP-----SPLKRAQPKQVPPKTSYLQM 170
QY	62 STSMKP 67
DB	171 PMMPP 176

RESULT 2

A59288
myosin heavy chain Myr 8 - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 09-Jun-2000 #sequence,revision 09-Jun-2000 #text_change 08-Sep-2000
C/Accession: A59288
R/Patel, K.G.; Liu, C.; Cameron, P.L.; Cameron, R.S.
Submitted to GenBank, November 1999
A/Description: Identification of a Novel Mammalian Myosin Class, XVI, in Developing
A/Reference number: A59288
A/Accession: A59288
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-1322 <PAT>
A/Cross-references: GB:AF209114; PIDN:AAF20150.1
C/Experimental source: strain Sprague-Dawley; clone KP4; cell type type 1 astrocyte
C/Superfamily: myosin motor domain homology <MMO>
E/404-1132/Domain: myosin motor domain homology <MMO>

Query Match 21.4%; Score 80.5; DB 2; Length 1322;
Best local similarity 38.4%; Pred. No. 7.4;


```

OY      14 HSSQSPSQAVTVNSQPSLMAVNSLTRSGL-----KTPPSLKPVVP--KPSFAPLS-TS 64
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1489 HGTAPGSEALKEVAQPSKAAH---RRGLPEPMVSKEDPKEQDTLPLAPPSSLPISDTS 1544

OY      65 MKP 67
      ||
Db      1545 PKP 1547

RESULT 11
T41308
hypothetical zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: T41308
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Wedler, H.; Wambutt, R.; Wedler, E.
submitted to the EMBL Data Library, March 1998
A:Reference number: 221986
A:Accession: T41308
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-867 <WOO>
A:Cross-references: EMBL:AL022245; PIDN:CAA18305.1; GSPDB:GN00068; SPDB:SPCC320.03
C:Experimental source: strain 972h-; cosmid c320
C:Genetics:
A:Gene: SPDB:SPCC320.03
A:Map position: 3
C:Superfamily: GAL4 zinc binuclear cluster homology
F:71-113/Domain: GAL4 zinc binuclear cluster homology <GL4>

Query Match      19.4%; Score 73; DB 2; Length 867;
Best Local Similarity 29.4%; Pred. No. 25;
Matches 20; Conservative 11; Mismatches 35; Indels 2; Gaps 1;

OY      2 PPAQRVDST--QVHSSQPSQAVTVNSKQPSLMAVNSLTRSGLKRTPSLKPDVPPKPSFA 59
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      329 PTVNDKRVSNVLPSTSPDSSVTTPVNSPAILNSITTSVPSGMSRHPMLNPPSTPEPSLG 388

OY      60 PLTSMKP 67
      | | : |
Db      389 VNPSPLRP 396

RESULT 12
S01955
hypothetical protein, 69K - turnip yellow mosaic virus
C:Species: turnip yellow mosaic virus, TYMV
C:Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 17-Mar-2000
C:Accession: S01955
R:Morch, M.D.; Boyer, J.C.; Haenni, A.L.
Nucleic Acids Res. 16, 6157-6173, 1988
A:Title: Overlapping open reading frames revealed by complete nucleotide sequencing of t
A:Reference number: S01955; MUID:88289359; PMID:3393388
A:Accession: S01955
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-628 <MO>
A:Cross-references: EMBL:X07441; NID:G62222; PIDN:CAA30321.1; PID:G62223
A:Note: the authors translated the codon ACG for residue 459 as U
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match      19.3%; Score 72.5; DB 2; Length 628;
Best Local Similarity 29.6%; Pred. No. 19;
Matches 21; Conservative 7; Mismatches 14; Indels 29; Gaps 3;

OY      2 PPAQRVDSTVHSSQPSQAVTVNSKQPSLMAVNSLTRSGLKRTP-----SLKPDV-PPKP 56
      | | | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      119 PPAQRQHSPLPILHTIRPS-----RRPHNHARRPDVPLSPV 154

OY      57 SFAPLSTSMKP 67

```

```

Db          155 DHGCVLTETKRP 165
               ||| |
RESULT 13
B84869
probable srf16 protein (Helianthus annuus) [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 17-May-2002
C:Accession: B84869
R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.;
M.; Koop, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MID:20083487; PMID:10617197
A:Accession: B84869
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-657 <STD>
A:Cross-references: GB:AEO02093; NID:g2281102; PIDN:AAE64038.I; GSPDB:GM00139
C:Genetics:
A:Gene: At2g43680
A:Map position: 2
C:Superfamily: Arabidopsis thaliana hypothetical protein Tr16L24.240

Query Match      19.3%; Score 72.5; DB 2; Length 657;
Best Local Similarity 35.1%; Pred. No. 20;
Matches 27; Conservative 9; Mismatches 28; Indels 13; Gaps 4;

OY      1 PPPAPQRDSTQVHSSOPSGQAVTVSRQPSLNAYNSLTRSGLKRTPKDPVP---PKP 56
           |||| :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      80 PPRPPRA--SRRVASPRPTSPRVASRPVSPPRA-EVRPTLSPPKPSPAEVPRSLSPKP 134
           |||| | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      57 ---SFAPLSTSMKPN 69
           : | | | | | |
DB     135 PSPRADLPRLSPKPF 151

RESULT 14
T03455
ALR protein - human
C:Species: Homo sapiens (man)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455
R:Pisrad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Y
oncogene 15, 549-560, 1997
A>Title: Structure and expression pattern of human ALR, a novel gene with strong hom
A:Reference number: Z14954; MUTID:97388474; PMID:9247308
A:Accession: T03455
A>Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-957 <PRA>
A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAAC51735.1; PID:g2358287
C:Genetics:
A:Gene: ALR
A:Map position: 12
C:Superfamily: human ALR protein
C:Keywords: alternative splicing

Query Match      19.3%; Score 72.5; DB 2; Length 4957;
Best Local Similarity 34.7%; Pred. No. 1.8e+02;
Matches 25; Conservative 6; Mismatches 24; Indels 17; Gaps 4;

OY      12 QVHSOPSQGAVTVSRQPSLNAYNSLTRSGLKRT-----PSLRPDVP----PKP----- 56
           ::||| |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB     1265 ELHAKEVPGQPENFEFRSPGTGAIVG-TSPMRFTFPQAAGEPSSLKPPVFQGLPPPHGIN 1983
           |||| | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      57 --SFAPLSTSMKP 67
           ||| | | | | |
DB     1984 SHFGPGPTUGKP 1995

```

RESULT 15

T03454

ALR protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000

C:Accession: T03454

R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,

Oncogene 15; 549-560, 1997

A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog

A:Reference number: Z14954; MUID:97388474; PMID:9247308

A:Accession: T03454

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-5262 <PRA>

A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285

C:Genetics:

A:Gene: ALR

A:Map position: 12

C:Superfamily: human ALR protein

C:Keywords: alternative splicing

Query Match

Best Local Similarity 19.3%; Score 72.5; DB 2; Length 5262;

Matches 25; Conservative 6; Mismatches 24; Indels 17; Gaps 4;

QY 12 QVHSSQPSGQAVTVSROPISLNAVNSLRSGLKRT-----PSLKPDPV-----PKP----- 56

Db 2230 ELHAKVPSGQPPNFEVRSPTGAFVG-TPSPMRFTFPQAVGEPSLKPVPVQPGLPDPHGIN 2288

QY 57 -SFAPLSTSMKP 67

Db 2289 SHFGPGPTLGRP 2300

Search completed: September 30, 2003, 16:35:30

Job time : 55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 30, 2003, 16:32:18 ; Search time 445 Seconds
(without alignments)
24.481 Million cell updates/sec

Title: US-09-856-681-4

Perfect score: 1 PPAAPRVDSIQVHSSQPSG.....PPKPSFAPLSTSMKPNDACT 72

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 566894 seqs, 151307093 residues

Total number of hits satisfying chosen parameters: 566894

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	376	100.0	1047	11	US-09-957-187-85
2	86	22.9	873	10	US-09-843-245-3
3	80.5	21.4	374	12	US-10-239-431A-20
4	80.5	21.4	564	12	US-10-239-431A-8
5	74	19.7	628	15	US-10-213-980-36
6	74	19.7	884	15	US-10-303-683-21
7	74	19.7	894	11	US-09-291-417-18
8	74	19.7	894	15	US-10-303-683-20
9	73.5	19.5	2655	5	US-09-864-761-34248
10	73.5	19.5	3664	15	US-10-177-293-423
11	73	19.4	200	12	US-10-259-165-300
12	73	19.4	4019	10	US-09-738-973-425
13	73	19.4	4019	10	US-09-854-133-425
14	73	19.4	4019	15	US-10-144-649A-425
15	71.5	19.0	223	9	US-09-864-761-33417

16	71.5	19.0	360	9	US-09-925-301-1116
17	71.5	19.0	589	9	US-09-972-086-2
18	69.5	18.5	1125	15	US-10-100-957A-152
19	69.5	18.5	1610	15	US-10-100-957A-22
20	68.5	18.2	625	10	US-09-801-368-294
21	68.5	18.2	729	15	US-10-145-396-11
22	68	18.1	153	10	US-09-925-300-1631
23	68	18.1	602	11	US-09-934-455-122
24	68	18.1	697	15	US-10-218-957-4
25	68	18.1	812	15	US-10-218-957-2
26	68	18.1	1149	10	US-09-969-528-5
27	68	18.1	1258	10	US-09-982-543-1
28	67.5	18.0	274	15	US-10-188-702A-6
29	67.5	18.0	405	15	US-10-153-668-218
30	67.5	18.0	5179	9	US-09-922-217-1068
31	67.5	18.0	5179	10	US-09-833-263-1068
32	67.5	18.0	5179	14	US-10-025-380-1068
33	66.5	17.7	503	14	US-10-078-547-2
34	66.5	17.7	507	14	US-10-078-547-24
35	66	17.6	200	14	US-10-062-254-170
36	66	17.6	448	9	US-09-864-761-44230
37	66	17.6	479	9	US-09-925-302-527
38	65.5	17.4	144	12	US-10-263-828-77
39	65.5	17.4	187	14	US-10-062-254-168
40	65.5	17.4	261	14	US-10-062-254-140
41	65.5	17.4	261	14	US-10-062-254-142
42	65.5	17.4	276	9	US-09-864-761-38306
43	65.5	17.4	463	14	US-10-029-180-80
44	65.5	17.4	473	9	US-09-864-761-38321
45	65.5	17.4	1142	10	US-09-899-651-2

ALIGNMENTS

RESULT 1
US-09-957-187-85
Sequence 85, Application US/09957187
Publication No. US2003005451A1
GENERAL INFORMATION:
APPLICANT: Shimkels, Richard A.
TITLE OF INVENTION: NOVEL POLYPEPTIDES AND PROTEINS ENCODED THEREBY
FILE REFERENCE: 15966-540 CIP
CURRENT APPLICATION NUMBER: US/09/957,187
CURRENT FILING DATE: 2000-09-19
PRIOR APPLICATION NUMBER: 60/123,667
PRIOR FILING DATE: 1999-03-09
PRIOR APPLICATION NUMBER: 60/234,082
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: 60/233,798
PRIOR FILING DATE: 2000-09-19
PRIOR APPLICATION NUMBER: 60/174,485
PRIOR FILING DATE: 2000-01-04
NUMBER OF SEQ ID NOS: 85
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 85
LENGTH: 1047
TYPE: PRT
ORGANISM: Homo sapiens
US-09-957-187-85
Query Match 100.0%; Score 376; DB 11; Length 1047;
Best Local Similarity 100.0%; Pred. No. 1.9e-29;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAAPRVDSIQVHSSQPSGAVTVRQPSLNAVNSLRTSGLRKTPSLKRDVPPKPSFAP 60
|||||
DB 976 PPAAPRVDSIQVHSSQPSGAVTVRQPSLNAVNSLRTSGLRKTPSLKRDVPPKPSFAP 1035
QY 61 LSTSMKPNDACT 72

Db 1036 LSTSMKPNDACT 1047

RESULT 2

US-09-843-245-3
; Sequence 3, Application US/09843245
; Patent No. US20020164672A1
; GENERAL INFORMATION:
; APPLICANT: McPherson, Peter S.
; APPLICANT: Ramjaun, Antoine Rachid
; TITLE OF INVENTION: REGULATION OF JNK ACTIVITY BY MODULATION OF THE
; TITLE OF INVENTION: INTERACTION BETWEEN THE ENDOCYTIC PROTEIN ENDOPHILIN
; TITLE OF INVENTION: AND THE GERMINAL CENTER KINASE-LIKE KINASE
; FILE REFERENCE: 9555.1160S01
; CURRENT APPLICATION NUMBER: US/09/843.245
; CURRENT FILING DATE: 2001-04-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-843-245-3

Query Match 22.9%; Score 86; DB 10; Length 873;
Best Local Similarity 34.7%; Pred. No. 2;
Matches 25; Conservative 6; Mismatches 27; Indels 14; Gaps 3;

QY 1 PPPAPQVDSIQV-----HSSQPSGOAVTVSRQPSLMAVNSLTRSGIKRTPSLKPDVPPKPF 56
DB 411 PPPAPPKPKSISIPQDTHSHSDSNQG-TIKRCPS-----SGSPAKPSHVPPRPPPP 460
QY 57 SFAPLSTSMKPN 68
DB 461 RLPPQKPAVLGN 472

RESULT 3

US-10-239-431A-20
; Sequence 20, Application US/10239431A
; Publication No. US20030170726A1
; GENERAL INFORMATION:
; APPLICANT: FRADELIZE, JULIE
; APPLICANT: FRIEDERICH, EVELYNE
; APPLICANT: GOLSTEYN, ROY M.
; APPLICANT: LOUYARD, DANIEL
; APPLICANT: NOIREAUX, VINCENT
; APPLICANT: SYKES, CECILE
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
; TITLE OF INVENTION: TO PROTEINS OF THE Ema/VASP FAMILY, AND THEIR USES
; FILE REFERENCE: 0508-1032
; CURRENT APPLICATION NUMBER: US/10/239.431A
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/FR01/00843
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: FR 00/03637
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Murine sp.
US-10-239-431A-20

Query Match 21.4%; Score 80.5; DB 12; Length 374;
Best Local Similarity 32.8%; Pred. No. 2.6;
Matches 22; Conservative 11; Mismatches 25; Indels 9; Gaps 4;

QY 1 PPPAPQVDSIQVSSQPSGOAVTVSRQPSLMAVNSLTRSGIKRTPSLKPDVPPKPSFAP 60
DB 208 PPPQPAKPOVQULH-VQPOAKP-HVQPOP-VSSANTQPRGPLSQAPT-----PAKPFAP 258

QY 61 LSTSMKP 67
DB 259 VAPKFTP 265

RESULT 4

US-10-239-431A-8
; Sequence 8, Application US/10239431A
; Publication No. US20030170726A1
; GENERAL INFORMATION:
; APPLICANT: FRADELIZE, JULIE
; APPLICANT: FRIEDERICH, EVELYNE
; APPLICANT: GOLSTEYN, ROY M.
; APPLICANT: LOUYARD, DANIEL
; APPLICANT: NOIREAUX, VINCENT
; APPLICANT: SYKES, CECILE
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
; TITLE OF INVENTION: TO PROTEINS OF THE Ema/VASP FAMILY, AND THEIR USES
; FILE REFERENCE: 0508-1032
; CURRENT APPLICATION NUMBER: US/10/239.431A
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/FR01/00843
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: FR 00/03637
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Murine sp.
US-10-239-431A-8

Query Match 21.4%; Score 80.5; DB 12; Length 564;
Best Local Similarity 32.8%; Pred. No. 4.2;
Matches 22; Conservative 11; Mismatches 25; Indels 9; Gaps 4;

QY 1 PPPAPQVDSIQVHSSQPSGOAVTVSRQPSLMAVNSLTRSGIKRTPSLKPDVPPKPSFAP 60
DB 209 PPPQPAKPOVQULH-VQPOAKP-HVQPOP-VSSANTQPRGPLSQAPT-----PAKPFAP 259
QY 61 LSTSMKP 67
DB 260 VAPKFTP 266

RESULT 5

US-10-213-990-36
; Sequence 36, Application US/10213990
; Publication No. US20030082595A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Bussey, Howard
; APPLICANT: Storms, Reg
; APPLICANT: Roemer, Terry
; TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUNIGATUS ENCODING INDUSTRIAL
; TITLE OF INVENTION: ENZYMES AND METHODS OF USE
; FILE REFERENCE: 10182-019-999
; CURRENT APPLICATION NUMBER: US/10/213.990
; CURRENT FILING DATE: 2002-08-05
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 628
; TYPE: PRT
; ORGANISM: Aspergillus
US-10-213-990-36

Query Match 19.7%; Score 74; DB 15; Length 628;
Best Local Similarity 42.2%; Pred. No. 22;
Matches 19; Conservative 5; Mismatches 19; Indels 2; Gaps 1;

Best Local Similarity 30.4%; Pred. No. 7.3;
Matches 21; Conservative 11; Mismatches 33; Indels 4; Gaps 3;

QY 2 PPAPQVDSIOYHS-SQPSGQAVTVSRQPSLNAYNSLTRSGLRK-RTPSLKPDVPP-KPS 57
Db 129 PAVGAVQSQQISATSSPFAETSTCSRAAPRVYRTNSAGCRSRSPATPTSPAPSPS 188

QY 58 FAPLSTSMK 66

Db 189 AAPCASCRCR 197

RESULT 12

US-09-738-973-425
Sequence 425, Application US/09738973
Patent No. US20020110563A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Henderson, Robert A.

APPLICANT: Lodes, Michael J.

APPLICANT: Fling, Steven P.

APPLICANT: Mohamath, Raodoh

APPLICANT: Algate, Paul A.

APPLICANT: Secrist, Heather

APPLICANT: Inditias, Carol Joseph

APPLICANT: Benson, Darin R.

APPLICANT: Eliot, Mark

APPLICANT: Mannion, Jane

APPLICANT: Kalos, Michael D.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C9

CURRENT APPLICATION NUMBER: US/09/738,973

CURRENT FILING DATE: 2000-12-14

NUMBER OF SEQ ID NOS: 587

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 425

LENGTH: 4019

TYPE: PRT

ORGANISM: Homo sapiens

US-09-738-973-425

Query Match 19.4%; Score 73; DB 10; Length 4019;

Best Local Similarity 32.5%; Pred. No. 2.3e+02;

Matches 27; Conservative 11; Mismatches 27; Indels 18; Gaps 6;

QY 1 PPAPQV--DSI-QVHSSQPSGQAV-----TVSRQPS-LNAYNSLTRSGLRKP----- 46

Db 910 PPAPSRIPIDSLSQAGTSQPPSPQVFSFGSSNSRPPSPMDPYAKMV--GTPRPPVGH 967

QY 47 --SLKPDVPPKPSFAPLSTSMKP 67

Db 968 SFSRRNSAAPVENCPTLSSVSRP 990

RESULT 13

US-09-854-133-425
Sequence 425, Application US/09854133
Publication No. US20020183499A1

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Mohamath, Raodoh

APPLICANT: Henderson, Robert A.

APPLICANT: Benson, Darin R.

APPLICANT: Secrist, Heather

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C10

CURRENT APPLICATION NUMBER: US/09/854,133

CURRENT FILING DATE: 2001-05-11

NUMBER OF SEQ ID NOS: 735

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 425

LENGTH: 4019

TYPE: PRT

ORGANISM: Homo sapiens

US-09-854-133-425

Query Match 19.4%; Score 73; DB 10; Length 4019;

Best Local Similarity 32.5%; Pred. No. 2.3e+02;

Matches 27; Conservative 11; Mismatches 27; Indels 18; Gaps 6;

QY 1 PPAPQV--DSI-QVHSSQPSGQAV-----TVSRQPS-LNAYNSLTRSGLRKP----- 46

Db 910 PPAPSRIPIDSLSQAGTSQPPSPQVFSFGSSNSRPPSPMDPYAKMV--GTPRPPVGH 967

QY 47 --SLKPDVPPKPSFAPLSTSMKP 67

Db 968 SFSRRNSAAPVENCPTLSSVSRP 990

RESULT 14

US-10-144-649A-425

Sequence 425, Application US/10144649A

Publication No. US20030118599A1

GENERAL INFORMATION:

APPLICANT: Lodes, Michael J.

APPLICANT: Wang, Tongtong

APPLICANT: Fan, Liqun

APPLICANT: Algate, Paul A.

APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.475C11

CURRENT APPLICATION NUMBER: US/10/144,649A

CURRENT FILING DATE: 2002-08-21

NUMBER OF SEQ ID NOS: 749

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 425

LENGTH: 4019

TYPE: PRT

ORGANISM: Homo sapiens

US-10-144-649A-425

Query Match 19.4%; Score 73; DB 15; Length 4019;

Best Local Similarity 32.5%; Pred. No. 2.3e+02;

Matches 27; Conservative 11; Mismatches 27; Indels 18; Gaps 6;

QY 1 PPAPQV--DSI-QVHSSQPSGQAV-----TVSRQPS-LNAYNSLTRSGLRKP----- 46

Db 910 PPAPSRIPIDSLSQAGTSQPPSPQVFSFGSSNSRPPSPMDPYAKMV--GTPRPPVGH 967

QY 47 --SLKPDVPPKPSFAPLSTSMKP 67

Db 968 SFSRRNSAAPVENCPTLSSVSRP 990

RESULT 15

US-09-864-761-33417

Sequence 33417, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

```

: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 33417
: LENGTH: 223
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC006581.16
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 13
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 41
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 23
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 25
: OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 28
: OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 46
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 17
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 71
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 29
: OTHER INFORMATION: EST_HUMAN HIT: AW959289.1, EVALU6 9.00e-15
: US-09-864-761-33417

Query Match      19.0%; Score 71.5; DB 9; Length 223;
Best Local Similarity 35.8%; Pred. No. 12;
Matches 24; Conservative 5; Mismatches 23; Indels 15; Gaps 3;

QY      2 PPAPRVDSIQVHSQPSGQAVTVSRQPSLNAYNSLTSGLKRP-SLKPDVPPKPSFAP 60
      1 11 : 11 11 11 : 11 11 : 11 11 : 11 11 : 11 11 :
DB      79 PGPQVPPPTQQVPPSQSQQAQITVDPMLQS-----SPLSLPPDAAPKP---P 124
      : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
QY      61 LSTSMKP 67
      : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
DB      125 IPIQSKP 131
```

Search completed: September 30, 2003, 16:43:58
Job time : 449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 30, 2003, 16:19:53 : Search time 69 Seconds

(Without alignments)
269.272 Million cell updates/sec

Title: US-09-856-681-4

Perfect score: 376
Sequence: 1 PPAAPQVDSIQVHSSQPSG.....PKRSPAPLSTSMKPNDACT 72

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	376	100.0	507	4	Q96T04
2	376	100.0	562	4	Q96S14
3	376	100.0	562	4	Q8NC49
4	376	100.0	574	4	Q96SM8
5	376	100.0	699	4	Q96SW4
6	376	100.0	1005	11	Q96O71
7	376	100.0	1030	4	Q9H2E6
8	376	100.0	1049	4	Q9P2H9
9	163.5	43.5	416	6	Q95KA6
10	163.5	43.5	451	4	Q9H9K4
11	163.5	43.5	464	4	Q9H9G5
12	163.5	43.5	998	4	Q8NFX6
13	163.5	43.5	1011	4	Q8NFX3
14	163.5	43.5	1017	4	Q8NFX5
15	163.5	43.5	1022	4	Q9P249
16	163.5	43.5	1073	4	Q8NFX4

17	86	22.9	862	11	Q924I2
18	80.5	21.4	508	11	Q8CD55
19	80.5	21.4	564	11	Q8CB00
20	80.5	21.4	876	5	Q9XZNS
21	80.5	21.4	1322	11	Q9OX10
22	80.5	21.4	1912	11	Q9ERC1
23	80	21.3	477	6	Q97600
24	78	20.7	1220	5	Q9GPF9
25	77.5	20.6	616	4	Q9H6K5
26	77.5	20.6	1111	10	Q9S2L9
27	77	20.5	144	10	Q8GVN3
28	77	20.5	175	10	Q9M1T6
29	77	20.5	175	10	Q8GVN7
30	77	20.5	960	11	Q921I2
31	76.5	20.3	744	10	Q65375
32	76	20.2	508	4	Q9NX29
33	76	20.2	508	4	Q76049
34	75.5	20.1	312	10	Q9S174
35	75.5	20.1	441	5	Q965J5
36	75.5	20.1	493	4	Q8NFN7
37	75.5	20.1	698	5	Q8MSL9
38	75.5	20.1	732	3	Q8J1Y5
39	75.5	20.1	735	5	Q9W3N8
40	75.5	20.1	745	5	Q9U484
41	75.5	20.1	745	5	Q9W164
42	75.5	20.1	1216	3	Q9C276
43	74.5	19.8	449	5	Q46062
44	74.5	19.8	653	5	Q8MR25
45	74.5	19.8	1788	5	Q9VE13

ALIGNMENTS

RESULT 1

ID	Q96T04	PRELIMINARY:	PRT:	507 AA.
AC	Q96T04;			
DT	01-DEC-2001 (TREMBLrel. 19, last sequence update)			
DE	01-DEC-2001 (TREMBLrel. 19, last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, last annotation update)			
DE	Hypothetical protein FL14533.			
OS	Homo sapiens (human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Wagatsuma M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Yamanoto J., Makematsu A., Nakamura Y., Nagahara K., Masuno Y.,			
RA	Ninomiya K., Iwayanagi T.,			
RT	"NEO human cDNA sequencing project."			
RL	Submitted (May-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AK027439; BAB5111.1; -			
DR	InterPro: IPR003659; Plexin-like.			
KW	SMART: SM00423; PSI: 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 507 AA: 55464 MW: 805678438C51B39 CRC64;			

Query Match 100.0%; Score 376; DB 4; Length 507;
Best Local Similarity 100.0%; Pred. No. 5.6e-33;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 PPAAPQVDSIQVHSSQPSGQAVTVSRQSLNAYNSLTSGLKRTSLKPDVPPKSPAP 60
DB	436 PPAAPQVDSIQVHSSQPSGQAVTVSRQSLNAYNSLTSGLKRTSLKPDVPPKSPAP 495
QY	61 LSTSMKPNDACT 72

```

DB      496 LSTSMKPNDACT 507

RESULT 2
Q96SY4      PRELIMINARY;      PRT:      562 AA.
AC  Q96SY4;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  Hypothetical protein FLJ14565.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA  Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA  Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA  Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA  Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA  Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA  Niimiya K., Iwayanagi T.;
RT  "NEO human cDNA sequencing project.";
RL  Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
DR  EMBL: AK024471; BAB55136.1; -;
DR  InterPro: IPR003659; Plexin-like.
DR  SMART: SM00423; PSI: 1.
KW  Hypothetical protein.
SQ  SEQUENCE 562 AA; 61313 MW; 6AB3685FAD1DD78A CRC64;

Query Match
Best Local Similarity 100.0%; Score 376; DB 4; Length 562;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGIKRTPSLKPPVPPKPSFAP 60
DB      491 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGIKRTPSLKPPVPPKPSFAP 550
QY      61 LSTSMKPNDACT 72
DB      551 LSTSMKPNDACT 562

RESULT 3
Q8NC49      PRELIMINARY;      PRT:      562 AA.
AC  Q8NC49;
DT  01-OCT-2002 (TREMBLrel. 22, Created)
DT  01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT  01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE  Hypothetical protein FLJ90494.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Isegaki T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA  Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA  Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA  Masuno Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA  Hattori A., Okumura K., Iwayanagi T., Niimiya K.;
RT  "NEO human cDNA sequencing project.";
RL  Submitted (Mar-2002) to the EMBL/Genbank/DBJ databases.
DR  EMBL: AK074975; BAC11326.1; -;
KW  Hypothetical protein.
SQ  SEQUENCE 562 AA; 61286 MW; 708041459E34D78A CRC64;

Query Match
Best Local Similarity 100.0%; Score 376; DB 4; Length 562;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGIKRTPSLKPPVPPKPSFAP 60
DB      491 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGIKRTPSLKPPVPPKPSFAP 550
QY      61 LSTSMKPNDACT 72
DB      551 LSTSMKPNDACT 562

RESULT 4
Q96SM8      PRELIMINARY;      PRT:      574 AA.
AC  Q96SM8;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  Hypothetical protein FLJ14748.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA  Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA  Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA  Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA  Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA  Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA  Niimiya K., Iwayanagi T.;
RT  "NEO human cDNA sequencing project.";
RL  Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
DR  EMBL: AK02654; BAB55269.1; -;
DR  InterPro: IPR003659; Plexin-like.
DR  SMART: SM00423; PSI: 1.
KW  Hypothetical protein.
SQ  SEQUENCE 574 AA; 62822 MW; 0C79E01A4117A495 CRC64;

Query Match
Best Local Similarity 100.0%; Score 376; DB 4; Length 574;
Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGIKRTPSLKPPVPPKPSFAP 60
DB      503 PPPAPQRYDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGIKRTPSLKPPVPPKPSFAP 562
QY      61 LSTSMKPNDACT 72
DB      563 LSTSMKPNDACT 574

RESULT 5
Q96SM4      PRELIMINARY;      PRT:      699 AA.
AC  Q96SM4;
DT  01-DEC-2001 (TREMBLrel. 19, Created)
DT  01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT  01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE  Hypothetical protein FLJ14595.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_Taxid=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA  Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA  Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA  Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA  Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA  Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA  Niimiya K., Iwayanagi T.;

```

RT "NEDO human cDNA sequencing project."
 RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AA027501; BAB5158.1; -
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF01403; Sema; 1.
 DR SMART: SM00423; PSI; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 699 AA; 76723 MW; 2E5F11D59741394 CRC64;

Query Match 100.0%; Score 376; DB 4; Length 699;
 Best Local Similarity 100.0%; Pred. No. 8.1e-33;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 60
 DB 628 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 687
 QY 61 LSTSMKPNDACT 72
 DB 688 LSTSMKPNDACT 699

RESULT 6
 Q9EO71 PRELIMINARY; PRT; 1005 AA.

AC Q9EO71;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Axon guidance signal SEMA6A.
 GN SEMA6A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-20564339; PubMed-10993894;
 RA Klosternann A., Lutz B., Gertler F., Behl C.;
 RT "The orthologous human and murine semaphorin 6A-1 proteins (SEMA6A-1 like protein (EVL) via a novel catboxyl-terminal Zyxin-like domain.";
 RT J. Biol. Chem. 275:39647-39653(2000).
 RL EMBL: AF288666; AAG29494.1; -
 DR MGD: MGI:1203727; Sema6a.
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF01403; Sema; 1.
 DR SMART: SM00423; PSI; 1.
 DR SMART: SM00630; Sema; 1.
 SQ SEQUENCE 1005 AA; 111758 MW; 57B69927F45B079D CRC64;

Query Match 100.0%; Score 376; DB 11; Length 1005;
 Best Local Similarity 100.0%; Pred. No. 1.2e-32;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 60
 DB 934 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 993
 QY 61 LSTSMKPNDACT 72
 DB 994 LSTSMKPNDACT 1005

RESULT 7
 Q9H2E6 PRELIMINARY; PRT; 1030 AA.

AC Q9H2E6;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Semaphorin SEMA6A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20564339; PubMed-10993894;
 RA Klosternann A., Lutz B., Gertler F., Behl C.;
 RT "The orthologous human and murine semaphorin 6A-1 proteins (SEMA6A-1 like protein (EVL) via a novel catboxyl-terminal Zyxin-like domain.";
 RT J. Biol. Chem. 275:39647-39653(2000).
 RL EMBL: AF279656; AAG29378.1; -
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF01403; Sema; 1.
 DR SMART: SM00423; PSI; 1.
 DR SMART: SM00630; Sema; 1.
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 1049 AA; 116511 MW; 7781D20ACC7A8ABA CRC64;

Query Match 100.0%; Score 376; DB 4; Length 1030;
 Best Local Similarity 100.0%; Pred. No. 1.3e-32;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 60
 DB 959 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 1018
 QY 61 LSTSMKPNDACT 72
 DB 1019 LSTSMKPNDACT 1030

RESULT 8
 Q9P2H9 PRELIMINARY; PRT; 1049 AA.

AC Q9P2H9;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein KIAA1368 (Fragment).
 GN KIAA1368.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE-20181126; PubMed-10718198;
 RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.";
 RT DNA Res. 7:65-73(2000).
 RL EMBL: AB037789; BAA92606.1; -
 DR Genew: HGNC:10738; SEMA6A.
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR001627; Sema.
 DR Pfam: PF01403; Sema; 1.
 DR SMART: SM00423; PSI; 1.
 DR SMART: SM00630; Sema; 1.
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 1049 AA; 116511 MW; 7781D20ACC7A8ABA CRC64;

Query Match 100.0%; Score 376; DB 4; Length 1049;
 Best Local Similarity 100.0%; Pred. No. 1.3e-32;
 Matches 72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPAPQRVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLTRSGLKRTPSLKPDVPPKPSFAP 60

Db 978 PPAPQVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLT-----RSGLKRTPSLKPDVPPKPSFAP 1037
 QY 61 LSTSMKPNDACT 72
 Db 1038 LSTSMKPNDACT 1049

RESULT 9

Q95KA6 PRELIMINARY; PRT; 416 AA.

AC Q95KA6; PRELIMINARY; PRT; 416 AA.
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DE Hypothetical 45.8 kDa protein.
 OS Macaca fascicularis (Craeb eating macaque) (Cynomolpus monkey).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Medulla oblongata;
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
 RA Suzuki Y., Sugano S., Hashimoto K.;
 RT "Isolation of full-length cDNA clones from macaque brain cDNA
 libraries";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB063027; BAB6070.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 416 AA; 45771 MW; C84BE67EC2F69E2B CRC64;

Query Match 43.5%; Score 163.5; DB 6; Length 416;
 Best Local Similarity 50.7%; Pred. No. 8; 6e-10;
 Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;

QY 1 PPAPQVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLT-----RSGLKRTPSLKPDVPP 54
 Db 345 PPPTGAKVDYIO-----GTPVSYHLQPSLSRQSSYTSNGTLPTGKRTPSLKPDVPP 397
 QY 55 KPSFAPLSTSMKP 67
 Db 398 KPSFVQPTPSVRP 410

RESULT 10

Q9H9K4 PRELIMINARY; PRT; 451 AA.

AC Q9H9K4; PRELIMINARY; PRT; 451 AA.
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
 DE Hypothetical protein FLJ12685.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.;
 RT "NEO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK022747; BAB1422.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 451 AA; 49681 MW; EABBFEEF7067AB04 CRC64;

Query Match 43.5%; Score 163.5; DB 4; Length 451;
 Best Local Similarity 50.7%; Pred. No. 9; 4e-10;
 Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;

QY 1 PPAPQVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLT-----RSGLKRTPSLKPDVPP 54
 Db 380 PPPTGAKVDYIO-----GTPVSYHLQPSLSRQSSYTSNGTLPTGKRTPSLKPDVPP 432

QY 55 KPSFAPLSTSMKP 67
 Db 433 KPSFVQPTPSVRP 445

RESULT 11

Q9H9G5 PRELIMINARY; PRT; 464 AA.

AC Q9H9G5; PRELIMINARY; PRT; 464 AA.
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
 DE Hypothetical protein FLJ12769.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.;
 RT "NEO human cDNA sequencing project";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK022831; BAB1426.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 464 AA; 51214 MW; C850600BAE9A0C94 CRC64;

Query Match 43.5%; Score 163.5; DB 4; Length 464;
 Best Local Similarity 50.7%; Pred. No. 9; 7e-10;
 Matches 37; Conservative 8; Mismatches 15; Indels 13; Gaps 2;

QY 1 PPAPQVDSIQVHSSQPSGQAVTVSRQPSLNAYNSLT-----RSGLKRTPSLKPDVPP 54
 Db 393 PPPTGAKVDYIO-----GTPVSYHLQPSLSRQSSYTSNGTLPTGKRTPSLKPDVPP 445

QY 55 KPSFAPLSTSMKP 67
 Db 446 KPSFVQPTPSVRP 458

RESULT 12

Q8NFY6 PRELIMINARY; PRT; 998 AA.

AC Q8NFY6; PRELIMINARY; PRT; 998 AA.
 DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Semaphorin 6D isoform 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
 RA Qu X., Zhai Y., Wei H., Yu Y., Tang F., He F.;
 RT "Homo sapiens semaphorin 6D isoform 2 (SEMA6D.2) mRNA, complete cds.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF389427; AAM69450.1; -
 DR InterPro; IPR001627; Sema.
 DR Pfam; PF01403; Sema; 1.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 30, 2003, 15:15:52; Search time 38 Seconds

(without alignments)
(89,103 Million cell updates/sec)

Title: US-09-856-681-4

Perfect score: 376
Sequence: 1 PPPAPQVRVDSIQVHSSQPSG.....PPKPSFAPLSTSMKPNDACT 72

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SWISSPROT_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	376	100.0	1030	1	SM6A_HUMAN
2	87	23.1	961	1	PGD1_HUMAN
3	86.5	23.0	564	1	ZYX_MOUSE
4	86	22.9	862	1	MAK3_RAT
5	77	20.5	960	1	FGD1_MOUSE
6	75.5	20.1	397	1	GAT5_HUMAN
7	75.5	20.1	477	1	MA2_HUMAN
8	75.5	20.1	5147	1	PCLO_HUMAN
9	74	19.7	452	1	HIS7_PHYR
10	74	19.7	894	1	MAK3_HUMAN
11	73.5	19.5	5085	1	PCLO_RAT
12	72.5	19.3	628	1	V7OK_TYMW
13	71.5	19.0	446	1	TEB3_MOUSE
14	71.5	19.0	668	1	SCBL_HUMAN
15	71.5	19.0	1004	1	PHC1_HUMAN
16	71.5	19.0	1522	1	PSL1_SCHPO
17	71	18.9	428	1	EKL1_HUMAN
18	70.5	18.8	344	1	ZIPA_SHEON
19	70	18.6	719	1	PHL1_SCHPO
20	69.5	18.5	1012	1	PHC1_MOUSE
21	69.5	18.5	331	1	MA2_MESAU
22	69.5	18.5	477	1	MA2_MOUSE
23	69.5	18.5	1125	1	MAP4_MOUSE
24	68.5	18.2	625	1	R101_YEAST
25	68.5	18.2	743	1	TEB3_HUMAN
26	68	18.1	737	1	SKN1_CANAL
27	68	18.1	812	1	NAH2_HUMAN
28	68	18.1	1259	1	AUT2_HUMAN
29	67.5	18.0	425	1	EKL1_MOUSE
30	67.5	18.0	525	1	CO2A_HUMAN
31	67.5	18.0	628	1	V7OK_TYMW
32	67.5	18.0	628	1	V7OK_TYMW
33	67.5	18.0	5179	1	MOC2_HUMAN

34	67	17.8	315	1	YK04_CAEEL
35	67	17.8	529	1	DNB2_ADE05
36	67	17.8	638	1	KNC0_YEAST
37	67	17.8	779	1	SRP_DROME
38	67	17.8	813	1	MAH2_RAT
39	67	17.8	1078	1	S24A_HUMAN
40	67	17.8	1152	1	MAP4_HUMAN
41	67	17.8	2035	1	HFCL_HUMAN
42	66.5	17.7	503	1	MAIP_HUMAN
43	66.5	17.7	837	1	ROD1_YEAST
44	66	17.6	295	1	PRIB_MYCLE
45	66	17.6	433	1	HXB3_MOUSE

ALIGNMENTS

RESULT 1
SM6A_HUMAN
ID SM6A_HUMAN STANDARD; PRT; 1030 AA.
AC Q9H2E6; Q9P2H9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Semaphorin 6A precursor (Semaphorin VIA) (Sema VIA) (Semaphorin 6A-1) (SEMA6A-1).
GN SEMA6A OR KIAA1368.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP MEDLINE=20564339; PubMed=10993894;
RA Klosternann A., Lutz B., Gertler F., Behl C.;
RT "The orthologous human and murine semaphorin 6A-1 proteins (SEMA6A-1/Sema6A-1) bind to the enabled/Vasodilator-stimulated phosphoprotein-like protein (EVL) via a novel carboxyl-terminal tyrosine-like domain.";
RT J. Biol. Chem. 275:39647-39653(2000).
RL [2]
RN [2]
RP MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
CC -!- FUNCTION: Can act as repulsive axon guidance cues. May play a role in channeling sympathetic axons into the sympathetic chains and controlling the temporal sequence of sympathetic target innervation (By similarity).
CC -!- SUBUNIT: Active as a homodimer or oligomer. Interacts with EVL.
CC -!- SUPPLEMENTARY LOCATIONS: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: Named isoforms=2;
CC Name=1;
CC IsoId=Q9H2E6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9H2E6-2; Sequence=VSP_007113;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC -!- SIMILARITY: Contains 1 Sema domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL collaboration - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC


```

DR SMART; SM00325; RHOGEF; 1.
DR PROSITE; PS00741; DH_1; FALSE_NEG.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS50178; ZF_FIVE; 1.
KW Guanine-nucleotide releasing factor; zinc-finger; Repeat;
KW Disease mutation.
FT DOMAIN 373 561 DH.
FT DOMAIN 7 330 PRO-RICH.
FT SITE 171 179 SH3-BINDING (POTENTIAL).
FT SITE 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 590 689 PH 1.
FT ZN_FING 730 790 FYVE-TYPE.
FT DOMAIN 821 921 PH 2.
FT VARIANT 522 522 R -> H (IN AAS).
FT VARIANT 610 610 /FTID=VAR_015236.
FT VARIANT 610 610 /FTID=VAR_015237.
FT CONFLICT 10 23 AGPSEPEHPATNP -> RRAFGARTPGHEPA (IN REF.
FT CONFLICT 10 23 1).
FT CONFLICT 195 195 A -> G (IN REF. 1).
SQ SEQUENCE 961 AA; 106560 MW; 30963F7B931E45C CRC64;

Query Match
Best local Similarity 23.1%; Score 87; DB 1; Length 961;
Matches 23; Conservative 5; Mismatches 22; Indels 16; Gaps 2;

QY 2 PPAPORVDSIQVHSSQSGQAVTVSRQPSLNAYNSLTSGIKRTPSLKPDVPPKSPFAPL 61
DB 127 PEGPQRL-----KSDPPPTETTPSQR-----SPLKRAPGPKQVPPKPSYLQM 170
QY 62 STSMKP 67
DB 171 PEMPP 176

RESULT 3
ZYX_MOUSE
ID ZYX_MOUSE STANDARD; PRT; 564 AA.
AC 062523; P70461;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zyxin.
GN ZYX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR X Swiss Webster;
RX MEDLINE=97094926; PubMed=8940160;
RA Macalima T., Ote J., Henslik M.E., Bockholt S.M., Louis H.A.,
RA Kalif-Suske M., Grzeschik R.H., von der Ahe D., Beckerle M.C.;
RT "Molecular characterization of human zyxin."
RT J. Biol. Chem. 271:31470-31478(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ote J., Heischmann A., Breier G., Beckerle M.C., von der Ahe D.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ADHESION PLAQUE PROTEIN. BINDS ALPHA-ACTININ AND THE CRP
CC PROTEIN. MAY BE A COMPONENT OF A SIGNAL TRANSDUCTION PATHWAY THAT
CC MEDIATES ADHESION-STIMULATED CHANGES IN GENE EXPRESSION (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOSOL; ASSOCIATES WITH THE ACTIN
CC CYTOSKELETON NEAR THE ADHESION PLASMA.
CC -1- SIMILARITY: Contains 3 LIM zinc-binding domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y07711; CA68984.1; -
DR EMBL; X99063; CA67510.1; -
DR MGD; MGI:103072; Zyx.
DR InterPro; IPR001781; LIM.
DR Pfam; PF00412; LIM; 3.
DR ProDom; PD000094; LIM; 3.
DR SMART; SM00132; LIM; 3.
DR PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR PROSITE; PS50023; LIM_DOMAIN_2; 3.
KW Repeat; LIM domain; Metal-binding; Zinc; Cell adhesion.
FT DOMAIN 64 77 PRO-RICH.
FT DOMAIN 94 138 PRO-RICH.
FT DOMAIN 376 435 LIM 1.
FT DOMAIN 436 495 LIM 2.
FT DOMAIN 496 562 LIM 3.
FT CONFLICT 215 215 R -> A (IN REF. 1).
FT CONFLICT 284 282 IKKCLRMP -> NQKAVPPDA (IN REF. 1).
SQ SEQUENCE 564 AA; 60790 MW; 001E1B3C82ADALEB CRC64;

Query Match
Best local Similarity 23.0%; Score 86.5; DB 1; Length 564;
Matches 23; Conservative 11; Mismatches 24; Indels 9; Gaps 4;

QY 1 PPAPORVDSIQVHSSQSGQAVTVSRQPSLNAYNSLTSGIKRTPSLKPDVPPKSPFAPL 60
DB 209 PEPQFQKRPQVQVH-VQPAKP-HVQPCP-VSSANTQPPGLSQAPT-----PAKRPAP 259
QY 61 LSTSMKP 67
DB 260 VAPKFT 266

RESULT 4
MAK3_RAT
ID MAK3_RAT STANDARD; PRT; 862 AA.
AC 092412;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mitogen-activated protein kinase kinase kinase 3 (MEKKK 3)
DE (MARK/ERK kinase kinase 3) (MEK kinase kinase 3) (Fragment).
DE (Germinal center kinase related protein kinase) (GK) (Fragment).
GN MAPK3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH SH3GL2.
RX MEDLINE=21869947; PubMed=11384986;
RA Ramjaun A.R., Angers A., Legendre-Guillemin V., Tong X.-K.,
RA McPherson P.S.;
RT "Endophilin regulates JNK activation through its interaction with the
RT germinal center kinase-like kinase."
RL J. Biol. Chem. 276:28913-28919(2001).
CC -1- FUNCTION: May play a role in the response to environmental stress.
CC Appears to act upstream of the c-Jun N-terminal pathway (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- COFACTOR: Magnesium (By similarity).
CC -1- SUBUNIT: Interacts with SH3GL2. Interaction appears to regulate
CC MAPK3-mediated JNK activation.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -1- SIMILARITY: Contains 1 CNH domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

DR EMBL: AF312224; AAK53214.1; -
 DR HSSP: P24941; 1B0H.
 DR InterPro: IPR001180; Citron.
 DR InterPro: IPR000719; Prot_Kinase.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR Pfam: PF00780; CNH; 1.
 DR Pfam: PF00069; kinase; 1.
 DR Prodom: PD000001; Prot_kinase; 1.
 DR SMART: SM00036; CNH; 1.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
 KW ATP-binding; Transferase; Serine/threonine-protein kinase.
 FT NON_TER 1
 FT DOMAIN 5 262 PROTEIN_KINASE.
 FT NP_BIND 530 842 CNH.
 FT BINDING 11 19 ATP (BY SIMILARITY).
 FT BINDING 34 34 ATP (BY SIMILARITY).
 FT ACT_SITE 125 125 BY SIMILARITY.
 SQ SEQUENCE 862 AA; 97390 MW; 58013AC3B0A3287F CRC64;

Query Match 22.9%; Score 86; DB 1; Length 862;
 Best Local Similarity 34.7%; Pred. No. 0.72;
 Matches 25; Conservative 6; Mismatches 27; Indels 14; Gaps 3;

QY 1 PPAPOQVDSIOV-----HSSQPSGAQVTVSRPSLNAYNSLTRSGLKRPSTKPDVPPK 56
 DB 400 PEPAPRKXSIISIRPDTHSSDSDNG-TIKKCP-----SSPAKPSHVPPPPPP 449

OY 57 SPAPLSTSMKPN 68
 DB 450 RLPPQKPAVLGN 461

RESULT 5
 FGDL_MOUSE STANDARD; PRT; 960 AA.
 AC P52734;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative Rho/Rac guanine nucleotide exchange factor (Rho/Rac GEF)
 DE (Facio-genital dysplasia protein homolog).
 GN FGDL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96081343; PubMed=8535076;
 RA Pastoris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,
 RA Gorski J.L.;
 RT "Cloning and regional localization of the mouse facio-genital
 RT dysplasia (Fgdl) gene.";
 RL Mamm. Genome 6:658-661(1995).
 CC -1- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
 CC BY EXCHANGING BOUND GDP FOR FREE GTP.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: Contains 1 DBL-homology (DB) domain.
 CC -1- SIMILARITY: Contains 2 PH domains.
 CC -1- SIMILARITY: Contains 1 FYVE-type zinc finger.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).

DR EMBL: U22325; AAA96001.1; -
 DR HSSP: Q07889; IAME.
 DR MGD: MGI:104566; Fgdl.
 DR InterPro: IPR001331; GDS_CDC24.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000219; RhoGEF.
 DR InterPro: IPR000306; Znf_FYVE.
 DR Pfam: PF01363; FYVE; 1.
 DR Pfam: PF00169; PH; 2.
 DR Pfam: PF00621; RHOGEF; 1.
 DR SMART: SM00064; FYVE; 1.
 DR SMART: SM00233; PH; 2.
 DR SMART: SM00325; RHOGEF; 1.
 DR PROSITE: PS00741; DH_1; FALSE_NEG.
 DR PROSITE: PS50010; DH_2; 1.
 DR PROSITE: PS50003; PH_DOMAIN; 2.
 DR PROSITE: PS50178; ZF_FYVE; 1.
 KW Guanine-nucleotide releasing factor; Zinc-finger; Repeat.
 FT DOMAIN 372 560 DH.
 FT SITE 7 330 PRO-RICH.
 FT SITE 171 179 SH3-BINDING (POTENTIAL).
 FT SITE 179 187 SH3-BINDING (POTENTIAL).
 FT DOMAIN 589 688 PH 1.
 FT ZN_FING 729 789 FYVE-TYPE.
 FT DOMAIN 820 920 PH 2.
 SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 20.5%; Score 77; DB 1; Length 960;
 Best Local Similarity 31.8%; Pred. No. 5.9;
 Matches 21; Conservative 5; Mismatches 24; Indels 16; Gaps 2;

QY 2 PPAPOQVDSIOVHSSQPSGAQVTVSRPSLNAYNSLTRSGLKRPSTKPDVPPKSPAPL 61
 DB 127 PEGPQRL-----RSDPPPEIPEIPRP-----SPLKRAPGPKPDVPPKSPYLGM 170

OY 62 STSMKP 67
 DB 171 PVLPLP 176

RESULT 6
 GAT5_HUMAN STANDARD; PRT; 397 AA.
 AC Q9BMX5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription factor GATA-5 (GATA binding factor-5).
 GN GAT5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21636749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Cordy N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levesajaino M.H., Levesajaino M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.L.,
 RA Phillimore B.J.C.T., Pratchallam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilmshurst L., Wray P.W., Hubbard T., Durbini R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 CC -1- FUNCTION: BINDS TO THE FUNCTIONALLY IMPORTANT CEF-1 NUCLEAR
 CC PROTEIN BINDING SITE IN THE CARDIAC-SPECIFIC SLOW/CARDIAC TROPONIN
 CC C TRANSCRIPTIONAL ENHANCER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC TRANSCRIPTIONAL PROGRAM(S) THAT UNDERLIES SMOOTH MUSCLE CELL
 CC DIVERSITY (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -1- SIMILARITY: Contains 2 GATA-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AL499627; CAC36001.1; -;
 CC DR HSSP: P17679; IGFN.
 CC DR Genew: HGNC:15802; GATA5.
 CC DR InterPro: IPR000679; Znf_GATA.
 CC DR Pfam: PF05349; GATA-N; 1.
 CC DR Pfam: PF05349; GATA-N; 1.
 CC DR PRINTS: PR00619; GATAZNFINGER.
 CC DR SMART: SM00401; Znf_GATA; 2.
 CC DR PROSITE: PS00344; GATA_ZN_FINGER_1; 2.
 CC DR PROSITE: PS00344; GATA_ZN_FINGER_2; 2.
 CC KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
 CC Nuclear protein.
 CC FT ZN_FING 189 213 GATA-TYPE.
 CC FT ZN_FING 243 267 GATA-TYPE.
 CC SQ SEQUENCE 397 AA; 41299 MW; 5DFBA02085695C57 CRC64;
 CC
 CC Query Match 20.1%; Score 75.5; DB 1; Length 397;
 CC Best Local Similarity 29.3%; Pred. No. 3;
 CC Matches 22; Conservative 8; Mismatches 36; Indels 9; Gaps 1;
 CC
 CC QY 1 PPPAPQVDSICVHSSOP-----SGQAVTSPKSLNLYNSLTRSGLRTPSLKRD 51
 CC Db 277 PPPPLAMKKEISITRRKRPRTAKAKSSSTRNMAASPSAVASTDSSAATSKAKPSLASP 336
 CC QY 52 VPKKPSFAPLSTSMK 66
 CC Db 337 VCPGSPMAPQASGOE 351
 CC
 CC MAZ_HUMAN STANDARD; PRT; 477 AA.
 CC AC P56270; Q15703; G99443;
 CC DT 15-JUL-1998 (Rel. 36) Created)
 CC DT 15-JUL-1998 (Rel. 36) Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42) Last annotation update)
 CC DE Myc-associated zinc finger protein (MAZI) (Purine-binding
 CC DE transcription factor) (Pur-1) (ZF87) (ZIF87).
 CC GN MAZ.
 CC OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92366479; PubMed=1502157;
 RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;
 RT "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences
 RL regulating transcriptional initiation and termination.";
 RN Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Carcinoma;
 RX MEDLINE=92232709; PubMed=1567856;
 RA Pyrc J.J., Moberg K.H., Hall D.J.;
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds
 RL to two sites within the c-myc promoter.";
 RN Biochemistry 31:4102-4110(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancratic islets;
 RX MEDLINE=96428591; PubMed=8831693;
 RA Tsutsui H., Sakatsune O., Itakura K., Yokoyama K.K.;
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human
 RL pancreatic islet cells.";
 RN Biochem. Biophys. Res. Commun. 226:801-809(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96224025; PubMed=8626793;
 RA Parks C.L., Shenk T.;
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that
 RL responds to MAZ and Sp1.";
 RN J. Biol. Chem. 271:4417-4430(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphoblastoma;
 RX MEDLINE=98352105; PubMed=9685418;
 RA Song J., Murakami H., Tsutsui H., Tang X., Matsumura M., Itakura K.,
 RA Kanazawa I., Sun K., Yokoyama K.K.;
 RT "Genomic organization and expression of a human gene for Myc-
 RL associated zinc finger protein (MAZ).";
 RN J. Biol. Chem. 273:20603-20614(1998).
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES
 CC IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,
 CC MEI1 AND MEI2, WITHIN THE C-MYC PROMOTER HAVING GREATER
 CC AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES
 CC WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL
 CC MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.
 CC -1- SIMILARITY: Contains 6 C2H2-type zinc fingers.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M94046; -; NOT_ANNOTATED_CDS.
 CC EMBL: M93339; -; NOT_ANNOTATED_CDS.
 CC EMBL: D85131; BAAL2728.1; ALT_INIT.
 CC DR EMBL: U33819; AAB04121.1; ALT_INIT.
 CC DR EMBL: AB017335; BAA33064.1; -;
 CC DR PIR: A42170; A42170.
 CC DR TRANSFAC: T00490; -;
 CC DR TRANSFAC: T02305; -;
 CC DR Genew: HGNC:6914; MAZ.
 CC DR MIM: 600999; -;
 CC GO: GO:0006367; P:transcription initiation from Pol II promoter; TAS.
 CC GO: GO:0006369; P:transcription termination from Pol II promoter; TAS.
 CC InterPro: IPR007087; ZnfC2H2.
 CC Pfam: PF00096; Zf-C2H2; 6.

DR Prodom: PD0000003; znf_C2H2; 1.
 DR SMART: SM00355; znf_C2H2; 6.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 5.
 KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;
 FM RNA-binding; Repeat; Nuclear protein.
 FT ZN_FING 190 212 C2H2-TYPE 1.
 FT ZN_FING 279 301 C2H2-TYPE 2.
 FT ZN_FING 307 329 C2H2-TYPE 3.
 FT ZN_FING 337 360 C2H2-TYPE 4.
 FT ZN_FING 366 388 C2H2-TYPE 5.
 FT ZN_FING 392 413 C2H2-TYPE 6.
 FT ZN_FING 96 108 POLY-ALA.
 FT DOMAIN 133 139 POLY-PRO.
 FT DOMAIN 157 161 POLY-ALA.
 FT DOMAIN 245 249 POLY-GLY.
 FT DOMAIN 435 449 POLY-ALA.
 FT CONFLICT 259 259 MISSING (IN REF. 3).
 FT CONFLICT 401 401 L->M (IN REF. 2 AND 4).
 FT CONFLICT 443 447 MISSING (IN REF. 3).
 SQ SEQUENCE 477 AA; 48607 MW; C04C80F32C3C6825 CRC64;

Query Match 20.1%; Score 75.5; DB 1; Length 477;
 Best Local Similarity 25.8%; Pred. No. 3.7;
 Matches 23; Conservative 15; Mismatches 32; Indels 19; Gaps 3;

QY 1 PPPAPQ-----RDSIOY-----HSSQSGCAVTVSRQPSLNAVSLRSGIK 43
 Db 69 PPTTPGAPAPAEPLVDLFLVLAAGDSAAAAAATAAATAAAPPAPAASTVDPAALK 128
 QY 44 RPPSLKPDVPPKPSFAPLSTSMKPNDACT 72
 Db 129 QPPA--PPPPPPVSAAPAAAPASAAAT 155

RESULT 8

PCLO_HUMAN

STANDARD: PRT: 5147 AA.

AC 09Y6V0: 043373; 060305; 09BYC8; 09UTV2; 09Y6V9;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE 15-SEP-2003 (Rel. 42, Last annotation update)

DE Piccolo protein (Aczonin) (Fragments).

GN PCLO OR ACZ OR KIAA0559.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-759 FROM N.A.

RC TISSUE-Brain;

RX MEDLINE-99439764; PubMed-10508862;

RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,

RA Killmann M.W.;

RT "Aczonin, a 550-kD putative scaffolding protein of presynaptic active

RT zones, shares homology regions with rim and bassoon and binds

RT profilin.";

RT J. Cell Biol. 147:151-162(1999).

RL [2]

RP SEQUENCE OF 552-4404 FROM N.A.

RA Kraemer J., Wollam C., Wohldmann P., McGrane B.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).

RC TISSUE-Brain;

RX MEDLINE-98290545; PubMed-9628581;

RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,

RA Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. IX.

RT The complete sequences of 100 new cDNA clones from brain which can

RT code for large proteins in vitro.";

RL DNA Res. 5:31-39(1998).

RN [4]

RP SEQUENCE OF 4405-4439 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE-22388257; PubMed-12477932;
 RA Strausberg R.L., Fellgold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cassavati T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huylk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shychenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP SEQUENCE OF 4405-5147 FROM N.A.

RA Kallunki J., Elliott G.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: May act as a scaffolding protein involved in the

CC organization of synaptic active zones and in synaptic vesicle

CC trafficking (By similarity).

CC -1- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).

CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of

CC synaptic junctions (By similarity).

CC -1- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=2;

CC Comment-Additional isoforms seem to exist;

CC Name=1;

CC Name=2;

CC IsoId=Q9Y6V0-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,

CC VSP_003926, VSP_003927;

CC Note-No experimental confirmation available;

CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and

CC phospholipids. Calcium binds with low affinity but with high

CC specificity and induces a large conformational change.

CC -1- SIMILARITY: Contains 2 C2 domains.

CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC EMBL: Y19188; CAB60727.1; -

CC EMBL: AC004903; RAND20936.1; -

CC EMBL: AC004886; RAD21789.1; -

CC EMBL: AB011131; BAA25485.1; -

CC EMBL: BC001304; AAH01304.1; -

CC EMBL: AC004082; AAB97937.1; -

CC PIR: T00634; T00634.

CC HSP: P04410; 1A25.

CC Gene: HGNC:13406; PCLO.

CC MIM: 604918; -

CC GO: 0005856; C:cytoskeleton; NAS.

CC GO: 0005202; C:synaptic junction; ISS.

CC GO: 0005509; F:calcium ion binding activity; ISS.

CC GO: 0005544; F:calcium-dependent phospholipid binding acti. . .; ISS.

CC GO: 0005522; F:profilin binding activity; ISS.

CC GO: 0007010; P:cytoskeleton organization and biogenesis; ISS.

CC GO: 0016080; P:synaptic vesicle targeting; ISS.

```

DR InterPro: IPR000008; C2.
DR InterPro: IPR001565; Synaptotagmin.
DR PRINTS: PR00360; C2DOMAIN.
DR PRINTS: PR00399; SYNAPTOTAGMIN.
DR SMART: SM00239; C2.
DR PROSITE: PS00499; C2_DOMAIN_1; 1.
DR PROSITE: PS00004; C2_DOMAIN_2; 2.
DR Calcium/phospholipid-binding; zinc; Metal-binding; zinc-finger;
KW Repeat; Alternative splicing.
FT 1
FT NON_TER 1
FT DOMAIN 400 465
FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-O-P-O-P-X.
FT ZN_FING 499 523
FT ZN_FING 969 992
FT 4-TYPE (POTENTIAL).
FT NON_CONS 1010 1011
FT 4-TYPE (POTENTIAL).
FT DOMAIN 2300 2325
FT 4391 4442
FT 4544 4633
FT 5031 5121
FT 4404 4404
FT C2 DOMAIN 1.
FT C2 DOMAIN 2.
FT S -> SCNGLGIRITGCKEIPGSGEIGAYIAKILPGSAAE
FT OTGLMEG (in isoform 2).
FT /FTID=VSP_003923.
FT K -> KPTDGETKVVSHPTGELIQ (in isoform 2).
FT /FTID=VSP_003924.
FT G -> GQYVVYQNAS (in isoform 2).
FT /FTID=VSP_003925.
FT /TAHKS -> SKRRK (in isoform 2).
FT /FTID=VSP_003926.
FT /MISSING (in isoform 2).
FT /FTID=VSP_003927.
SQ SEQUENCE 5147 AA; 563537 MW; CD5D84990498C3C CRC64;

Query Match 20.1%; Score 75.5; DB 1; Length 5147;
Best Local Similarity 31.4%; Pred. No. 53;
Matches 22; Conservative 5; Mismatches 18; Indels 25; Gaps 3;

OY 1 PEPAPQVDSTGVHSSQSGAVTVSRQPSLNAVNSLTRSGLKRPSPKPPKSPAP 60
DB 2378 PEPVPPKPSST-----PSGLVETHPREPS-----KPTAPKPVIPQ 2413
OY 61 L-STSKMPND 69
DB 2414 LPTTQKRPD 2423

RESULT 9
HIS7_PHYPR
ID HIS7_PHYPR STANDARD: PRT; 452 AA.
AC P28624;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imlidazoleglycerol-phosphate dehydratase (EC 4.2.1.19) (IGPD).
GN HIS3.
OS Eukaryophora parasitica (Potato buckeye rot agent).
OC Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC Phycophthora.
OX NCBL_TaxID=4792;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1829;
RA Balnusch-Weiter M., Karlovsky P., Prell H.H.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -i- CATALYTIC ACTIVITY: D-erythro-1-(imidazol-4-yl)glycerol 3-
CC phosphate = 3-(imidazol-4-yl)-2-oxopropyl phosphate + H(2)O.
CC -i- PATHWAY: Histidine biosynthesis; sixth step.
CC -i- SIMILARITY: Belongs to the imidazoleglycerol-phosphate dehydratase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z11591; CAA77675.1; -.
DR InterPro: IPR006438; HAD-SF-1A-hypl.
DR InterPro: IPR006543; Histidinol-phos.
DR InterPro: IPR005834; Hydrolase.
DR InterPro: IPR000807; IGPD.
DR Pfam: PF00702; Hydrolase; 1.
DR Pfam: PF00475; IGPD; 1.
DR PRODOM: PD002282; IGPD; 1.
DR TIGRPFAM: TIGR01548; HAD-SF-1A-hypl; 1.
DR TIGRPFAM: TIGR01656; Histidinol-ppas; 1.
DR PROSITE: PS00954; IGP_DEHYDRATASE_1; 1.
DR PROSITE: PS00955; IGP_DEHYDRATASE_2; 1.
KW Histidine biosynthesis; Lyase; Multifunctional enzyme.
FT 1
FT DOMAIN 234 452
FT IMIDAZOLEGLYCEROL-PHOSPHATE DEHYDRATASE.
SQ SEQUENCE 452 AA; 479661 MW; CA66BE32A9E53A1 CRC64;

Query Match 19.7%; Score 74; DB 1; Length 452;
Best Local Similarity 34.4%; Pred. No. 4.9;
Matches 21; Conservative 8; Mismatches 22; Indels 10; Gaps 3;

OY 12 QVHSQPSQAVTVSRQPSLNAVNSLTRSGLKRPSPKPPKSPAP 65
DB 111 ELHRQPKQMAVTVGR-PRKDCAKFLTHGIE--DLFVQIWLIEDCPKPSPEPTLLAL 166
OY 66 K 66
DB 167 K 167

RESULT 10
MAK3_HUMAN
ID MAK3_HUMAN STANDARD: PRT; 894 AA.
AC O81VH8; O81VH7; Q9UDM5; Q9Y6R5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mitogen-activated protein kinase kinase kinase 3 (EC 2.7.1.37)
DE (MAPK/ERK kinase kinase kinase 3) (MEKKK 3)
DE (Germinal center kinase related protein kinase) (GLK).
GN MAP4K3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBL_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), FUNCTION, TISSUE SPECIFICITY, AND
RP MOTAGENSIS OF IYS-48.
RC TISSUE=Macrophage, and Skeletal muscle;
RX MEDLINE=97420743; PubMed=9275185;
RA Diener K., Wang X.S., Chen C., Meyer C.F., Keesler G., Zukowski M.,
RA Tan T.-H., Yao Z.;
RT "Activation of the c-Jun N-terminal kinase pathway by a novel protein
RT kinase related to human germinal center kinase.";
RT Proc. Natl. Acad. Sci. U.S.A. 94:9687-9692(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RA Gorry M.C., Zhang Y., Marks J.J., Suppe B., Hart S., Cortell J.,
RA Pallos D., Hart T.C.;
RT "Physical/genetic map of the 2p22-2p21 region on chromosome 2.";
RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-712 FROM N.A. (ISOFORM 1).
RA Edwards J., Mohlmann P., Hawkins R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: May play a role in the response to environmental stress.
CC Appears to act upstream of the c-Jun N-terminal
CC pathway.
CC -i- CATALYTIC ACTIVITY: ATP + a protein = ADP + a

```


DR	EMBL:	AF445400;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445401;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445402;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445403;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445404;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445405;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445406;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445407;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445409;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445410;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445411;	AAAT75850.1;	JOINED.
DR	EMBL:	AF445412;	AAAT75850.1;	JOINED.
DR	EMBL:	AC007684;	AAF19240.1;	-
DR	HSSP:	P24941:	1B38.	
DR	GeneW:	HGNC:6665;	MARPK3.	
DR	MM:	604921;	-	
DR	InterPro:	IPR001180;	Citron.	
DR	InterPro:	IPR000719;	Prot_Kinase.	
DR	InterPro:	IPR002290;	Ser_thr_kinase.	
DR	InterPro:	IPR001245;	Tyr_kinase.	
DR	pfam:	PF00780;	CNH: 1	
DR	pfam:	PF00069;	Kinase; 1.	
DR	ProDom:	PD000001;	Prot_Kinase; 1.	
DR	SMART:	SMO0036;	CNH: 1.	
DR	SMART:	SMO0220;	S_TKC; 1.	
DR	PROSITE:	PS00107;	PROTEIN_KINASE_ATP; 1.	
DR	PROSITE:	PS50011;	PROTEIN_KINASE_DOM; 1.	
DR	PROSITE:	PS00108;	PROTEIN_KINASE_ST; FALSE NEG.	
DR	kw	AtP-binding; Transferase; Serine/threonine-protein kinase;		
DR	kw	Alternative splicing		
FT	DOMAIN	16	273	PROTEIN KINASE.
FT	DOMAIN	562	874	CNH.
FT	NP_BIND	22	30	ATP (BY SIMILARITY).
FT	BINDING	48	48	ATP.
FT	ACT_SITE	136	136	BY SIMILARITY.
FT	VASPLIC	1	12	MNPGDLSRRMP -> MA (In isoform 2).
FT	VASPLIC	352	372	/Fttd-VSP_007052.
FT	VARSPLIC			Missing (In isoform 3).
FT	MUTAGEN	48	48	/Fttd-VSP_007053.
FT	CONFLICT	392	392	K->E: LOSS OF KINASE ACTIVITY AND ABILITY TO ACTIVATE JNK FAMILY.
FT	SEQUENCE	894 AA;	101315 MW;	N -> D (IN REF. 1; AAC15472).
FT	SEQUENCE	894 AA;	101315 MW;	6EB77BBB345B733 CRC64;
QY	Query Match		19.7%;	Score 74; DB 1; Length 894;
Db	Best Local Similarity		32.8%;	Pred. No. 10;
Matches	21; Conservative		6; Mismatches	23; Indels 14; Gaps 3;
QY	1 PPAPQQRVDSI-----QVHSSPSQSGAVTVSOPSLMAVNSITFRGLKRTPTSLKQDVPPK 56	:: ::: :: :: ::	::	
Db	432 PEPPIPPKKSIPTIQEHMSTEDENQG-TIKRCP-----MSGSPAKPSQVPKPRPP 481			
QY	57 SFAP 60			
Db	482 RLPP 485			
RESULT 11				
PCLO_RAT	STANDARD;		PRT;	5085 AA.
AC	O9JRS6;	O9JRT1;		
DT	28-FEB-2003	(Rel. 41,	Created)	
DT	28-FEB-2003	(Rel. 41,	Last sequence update)	
DT	28-FEB-2003	(Rel. 41,	Last annotation update)	
DE	Piccolo protein	(Multidomain presynaptic cytomatrix protein).		
GN	PCLO.			
OS	Rattus norvegicus	(Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
CX	NCBI_TaxID=10116;			
NP	[1]			
NP	SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH PARAC1.			

Db	40	PPPAQVPEVIAKVQTHENPRYHLOQARROOVXOYLSTLGPKLASQALTP--PPGSS	97
Oy	59	A-PL 61	
Db	98	AQPL 101	
RESULT 14			
SCSEL_HUMAN			
ID	SCSEL_HUMAN	STANDARD;	PRT; 668 AA.
AC	095171;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Scellin.		
SCSEL			
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 443-457 AND 635-650.		
RC	TISSUE=ForeSkin;		
RX	MEDLINE=99030435; PubMed=9813070;		
RA	Champland M.-F., Burgess R.E., Jin W., Baden H.P., Olson P.F.;		
RT	"cDNA cloning and characterization of scellin, a LIM domain protein of the keratinocyte cornified envelope."		
RL	J. Biol. Chem. 273:31547-31554(1998).		
CC	-1- FUNCTION: May function in the assembly or regulation of proteins in the cornified envelope. The LIM domain may be involved in homotypic or heterotypic associations and may function to localize scellin to the cornified envelope.		
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic. May become cross-linked to membrane proteins by transglutaminase.		
CC	-1- TISSUE SPECIFICITY: Highly expressed in esophagus. It is also expressed in keratinocytes, amniotic tissue, foreskin stratum spinosum and stratum granulosum, hair follicle and nail.		
CC	-1- SIMILARITY: Contains 1 LIM zinc-binding domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AF045941; AAC78461.1; -.		
DR	Genew; HGNC:10573; SCSEL.		
DR	MIM; 604112; -.		
DR	GO; GO:0005737; C:cytoplasm; TAS.		
DR	GO; GO:0008544; P:epidermal differentiation; TAS.		
DR	InterPro; IPR001781; LIM.		
DR	ProDom; PD000094; LIM; 1.		
DR	SMART; SM00132; LIM; 1.		
DR	PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.		
DR	PROSITE; PS50023; LIM_DOMAIN_2; 1.		
DR	KW LIM domain; Metal-binding; Zinc; Repeat.		
FT	DOMAIN	599	665
FT			LIM
FT	REPEAT	231	543
FT		231	246
FT	REPEAT	247	266
FT		247	266
FT	REPEAT	267	286
FT		267	286
FT	REPEAT	287	306
FT		287	306
FT	REPEAT	307	326
FT		307	326
FT	REPEAT	327	346
FT		327	346
FT	REPEAT	347	366
FT		347	366
FT	REPEAT	367	386
FT		367	386
FT	REPEAT	387	406
FT		387	406
FT	REPEAT	407	426
FT		407	426
FT	REPEAT	427	445
FT		427	445
FT	REPEAT	446	464
FT		446	464
FT	REPEAT	466	486
FT		466	486
FT	REPEAT	486	506
FT		486	506
FT	REPEAT	506	526
FT		506	526
FT	REPEAT	526	546
FT		526	546
FT	REPEAT	546	566
FT		546	566
FT	REPEAT	566	586
FT		566	586
FT	REPEAT	586	606
FT		586	606
FT	REPEAT	606	626
FT		606	626
FT	REPEAT	626	646
FT		626	646
FT	REPEAT	646	666
FT		646	666
FT	REPEAT	666	686
FT		666	686
FT	REPEAT	686	706
FT		686	706
FT	REPEAT	706	726
FT		706	726
FT	REPEAT	726	746
FT		726	746
FT	REPEAT	746	766
FT		746	766
FT	REPEAT	766	786
FT		766	786
FT	REPEAT	786	806
FT		786	806
FT	REPEAT	806	826
FT		806	826
FT	REPEAT	826	846
FT		826	846
FT	REPEAT	846	866
FT		846	866
FT	REPEAT	866	886
FT		866	886
FT	REPEAT	886	

FT REPEAT 465 484 13.
 FT REPEAT 485 503 14.
 FT REPEAT 504 523 15.
 FT REPEAT 524 543 16.
 SQ SEQUENCE 668 AA: 75297 MW: 612BC368DBA9PB0 CRC64;

Query Match 19.0%; Score 71.5; DB 1; Length 668;
 Best Local Similarity 30.0%; Pred. No. 13;
 Matches 24; Conservative 11; Mismatches 20; Indels 25; Gaps 5;

OY 10 SIQVHSSQPSQ-----AVYSRQ---PSLNAYNSLTRSGL-KRTPSL 48
 DB 127 SLEVTIKIQGGSLNANTNTASTATTPVKKKRQSWPEPPPGYGNASSSTGTRRRRPGV 186
 OY 49 KPDVPPKPSFAPLSTSMKPN 68
 DB 167 HPIPPKPS-SPVSS---PN 202

RESULT 15

PHCL_HUMAN

ID P78364; OGMW3; Q9BU63; PRT; 1004 AA.

AC 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Polyhomeotic-like protein 1 (Early development regulator protein 1)

DE (HPH1).

GN PHC1 OR EDRI OR PH1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN MEDLINE=97220024; PUBMED=9121482;

RX Gunster M.J., Satijn D.P., Hamer K.M., den Blaauwen J.L.,

RT de Bruijn D., Alkema M.J., van Lohuizen M., van Driel R., Ote A.P.,

RT vertebrate polycomb-group protein Bmi1 and human homologs of

RT polyhomeotic.";

RT Mol. Cell. Biol. 17:2326-2335(1997).

RN [2]

RP SEQUENCE OF 504-1004 FROM N.A.

RC TISSUE=Lung, and Lymph;

RX MEDLINE=22388257; PubMed=12477932;

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raba S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huliy S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Keltman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,

RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -! FUNCTION: Involved in the regulation of Hox gene expression and

CC segment specification during paraxial mesoderm and neural crest

CC development (By similarity).

CC -! SUBUNIT: Homodimer. Interacts with Bmi1.

CC -! SUBCELLULAR LOCATION: Nuclear.

CC -! SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL; 089277; AAC51169.1; -
 DR EMBL; BC002871; AAH02871.1; ALT_INIT.
 DR EMBL; BC017748; AAH17748.1; ALT_INIT.
 DR Genew; HGNC:3182; PHC1.
 DR MIM: 602978; -
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0005515; P:protein binding activity; TAS.
 DR InterPro; IPR01660; SAM.
 DR Pfam; PF00536; SAM; 1.
 DR SMART; SM00454; SAM; 1.
 DR PROSITE; PS50105; SAM_DOMAIN; 1.
 DR Developmental protein; DNA-binding; Nuclear protein; Metal-binding;
 KW Zinc-finger.
 FT DOMAIN 425 441
 FT POLY-GLN.
 FT 940 1004
 FT SAM.
 FT ZN-FING 800 823
 FT C4-TYPE (POTENTIAL).
 FT CONFLICT 568 569 HL -> LK (IN REF. 1).
 FT CONFLICT 613 613 S -> T (IN REF. 1).
 FT CONFLICT 693 693 A -> G (IN REF. 1).
 FT CONFLICT 752 752 V -> G (IN REF. 1).
 FT CONFLICT 782 782 L -> F (IN REF. 1).
 FT CONFLICT 972 972 L -> F (IN REF. 1).
 SQ SEQUENCE 1004 AA: 105418 MW: 5E35765758904C4F CRC64;

Query Match 19.0%; Score 71.5; DB 1; Length 1004;
 Best Local Similarity 35.8%; Pred. No. 21;
 Matches 24; Conservative 5; Mismatches 23; Indels 15; Gaps 3;

OY 2 PPAPRVYSIOVHSSQPSQGVAVYVROPSLNAYNSLTRSGL-KRTP-SIKPDVPPKPSAP 60
 DB 448 PQPDVPPRTQGVPPESQSQQAQTLVYQMLGS-----SFLSPDAAPK---P 493
 OY 61 LSTSMKP 67
 DB 494 PIOSKP 500

Search completed: September 30, 2003, 16:33:00
 Job time : 40 secs